



**HAL**  
open science

## How Newspapers Portrayed COVID-19

Redwan Ahmed Rizvee, Moinul Zaber

► **To cite this version:**

Redwan Ahmed Rizvee, Moinul Zaber. How Newspapers Portrayed COVID-19. 1st International Conference on Computer Science Protecting Human Society Against Epidemics (ANTICOVID), Jun 2021, Virtual, Poland. pp.41-52, 10.1007/978-3-030-86582-5\_5 . hal-03746675

**HAL Id: hal-03746675**

**<https://inria.hal.science/hal-03746675>**

Submitted on 5 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

# How Newspapers Portrayed COVID-19\*

## A Study Based on United Kingdom and Bangladesh

Redwan Ahmed Rizvee<sup>1</sup>[0000-0001-5056-1701] and Moinul Zaber<sup>1,2,3</sup>

<sup>1</sup> Data and Design Lab ([dndlab.org](http://dndlab.org)), Dhaka, Bangladesh

<sup>2</sup> University of Dhaka, Bangladesh

<sup>3</sup> UNU-EGOV, United Nations University, Portugal  
[rizveeredwan.csedu@gmail.com](mailto:rizveeredwan.csedu@gmail.com), [zaber@unu.edu](mailto:zaber@unu.edu)

**Abstract.** In this study, we wanted to see how as a representative of a society newspaper portrayed COVID-19. For this purpose, this study considered two countries United Kingdom (UK) and Bangladesh (BD), and analyzed how COVID-19 as an external event was focused in the newspapers. To conduct the analysis, we handpicked a set of covid related feature terms, and using Latent Dirichlet Allocation (LDA) we verified the coherency of the chosen terms. The main finding of this study is that, initially as a new event COVID-19 found huge importance in the newspapers, but with gradual time progression, this became a new normal event and lost its initial insurgence of focus and took a stable condition. We also observed that despite being quite different demographically, geographically, economically along with being affected differently due to covid, UK and BD exhibit quite similar characteristics in portraying covid in newspapers. The decisions were arrived at by applying different types of Natural Language Processing (NLP) tools and statistical analysis. For experimentation, we collected all the published news articles from two newspapers, the Guardian and the Daily Star from January 1, 2020, to March 31, 2021.

**Keywords:** COVID-19 · Newspapers' portrayal · Media Response · Crisis communication · External Event effect · UK · Bangladesh.

## 1 Introduction

Due to severity and widespread WHO declared COVID-19 as a pandemic on March 2020<sup>4</sup> and the world is still not free from the grasp of the coronavirus. SARS-CoV-2 is the virus that causes COVID-19 and responsible for the respiratory illness symptoms<sup>5</sup>. Since the appearance of covid, the world has been at

---

\* This paper is a result of the project “SmartEGOV: Harnessing EGOV for Smart Governance (Foundations, methods, Tools) / NORTE-01-0145-FEDER-000037”, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (EFDR).

<sup>4</sup> [https://en.wikipedia.org/wiki/COVID-19\\_pandemic](https://en.wikipedia.org/wiki/COVID-19_pandemic)

<sup>5</sup> <https://tinyurl.com/4dva2zaf>

a stalemate and almost each country has been facing challenges in different sectors, e.g., health, education, economy, etc. Each country's government has taken strategies to stand against covid based on their capability and covid situation's severity, e.g., lockdown, vaccination, increasing tests, etc. But how the society is reacting is another important challenge, e.g., support towards govt.'s strategies, awareness towards covid's severity, most dominating public concerns, etc. As a representative of society media can be very useful to answer these questions.

Newspaper has always been a very popular and affordable medium to express society's concerns, status, morals, policies, etc and has been used as a good data source to conduct different types of opinion mining [9, ?] to understand media response. Many literature have used newspapers to understand various effects of covid in society [9, 9]. This study tried to find during this long period how COVID-19 was looked at in the newspaper. This study mainly focused on three important questions.

1. How society gave importance over covid.
2. Different types of portrayals of covid or which issues related to it were given more focus.
3. How covid's severity controlled media's. response

Based on economical, infrastructural, demographical, geographical and other internal parameters each country can be broadly categorized into two groups, developed and underdeveloped. Our main goal was to observe how the answers to the above-mentioned questions vary for both types of countries.

As representative of developed and underdeveloped countries, we considered UK<sup>6</sup> and BD<sup>7</sup> respectively. From different aspects, e.g., economy, infrastructure, demography, etc. these two countries exhibit opposite characteristics. Also, COVID-19 affected these countries way differently. Up to now (May 30, 2021)<sup>8</sup>, the total number of cases and deaths recorded in UK is around 4,480,945 and 127,775 whereas in Bangladesh is 797,386 and 12,549. In this study, we wanted to observe how UK and BD were affected due to covid and the associative response reflected in the newspaper along with a time series trend associativity between them. The main findings of this work based on the asked questions can be summarized as -

1. Covid as an external event was significantly highlighted initially. But with gradual time progression, the initial extreme highlight degraded and took almost a stable condition for both UK and BD.
2. Both for the developed (UK) and underdeveloped (BD) countries we found that similar issues related to covid were highly addressed.
3. The ratio of warning over comforting sentences was comparatively more pronounced and vivid in the charts for BD compared to UK in response to covid's insurgence.

<sup>6</sup> <https://tinyurl.com/44yh63v3>

<sup>7</sup> <https://data.worldbank.org/country/bangladesh>

<sup>8</sup> <https://tinyurl.com/9st4ys9h>

In the following sections, we will discuss our data collection and analysis procedure and validate our findings.

## 2 Background Study

As a representative of society newspaper has been widely used as a data source for opinion mining to understand different aspects of society. In [5], the authors used topic modeling approach on the collected news articles to extract opinions of people about ‘presidential election’. In [8], the authors used headlines to identify different types of sentiments through Support Vector Machine (SVM) based approaches. In [9], the authors showed how newspapers played a crucial role in development and adoption of different contract-tracing apps to fight against covid. In [10], the authors considered newspaper coverage to measure response to the covid19 crisis targeting New Zealand and Singapore. Their investigation showed that Singapore showed higher sentiment value compared to New Zealand in response to covid19 pandemic. They also showed the trends in mean media sentiment of different topics to understand their relative importance transition over their considered time period. So, newspapers have been used to analysis different aspects related to covid in literature. In this study, we wanted to observe the reflectance of covid in the newspapers within a time period from the context of developed (UK) and underdeveloped (BD) countries.

## 3 Data Description

As representative we considered two newspapers, the *Guardian*<sup>9</sup> and the *Daily Star*<sup>10</sup> for UK and BD respectively. Both of them are quite popular in their respective countries<sup>11,12</sup>. We collected all the news articles from January 1, 2020 to March 31, 2021. For each news article, we collected the headline, main news, URL, summary (if found), and publication date. For Guardian, we had around 14808 news articles comprising of 40625 unique words and for Daily Star, we had around 34995 news articles having 52775 unique words. The collected data can be found in this link<sup>13</sup>.

## 4 Data Analysis

In this section, we will analyze our collected data from different perspectives and measures.

---

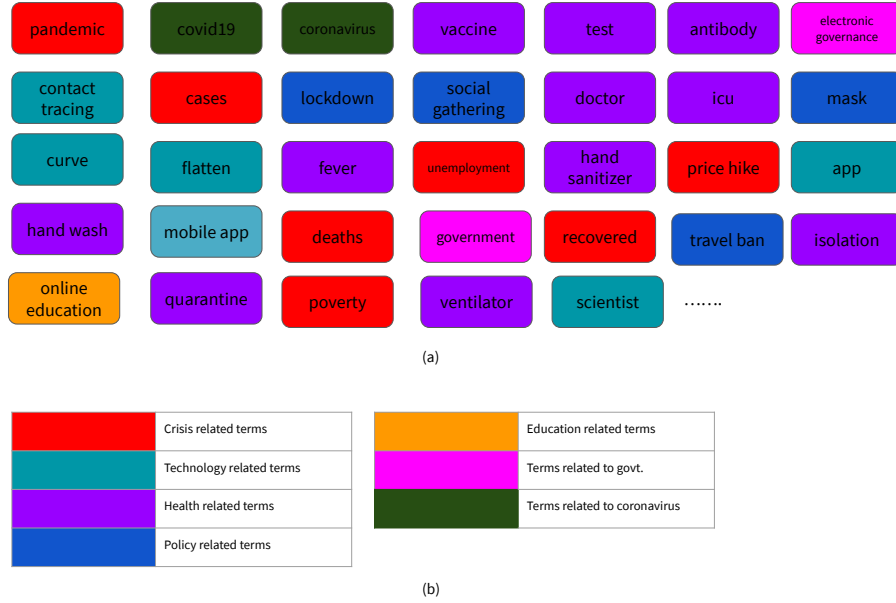
<sup>9</sup> <https://www.theguardian.com/>

<sup>10</sup> <https://www.thedailystar.net/>

<sup>11</sup> <https://tinyurl.com/25ufnb8n>

<sup>12</sup> <https://tinyurl.com/a87685ty>

<sup>13</sup> <https://tinyurl.com/w7y6kdvw>



**Fig. 1.** (a) Some of the handpicked features with their corresponding category, (b) Color Coding to represent the categories

#### 4.1 Validation of covid related handpicked features

We have constructed a set of popular covid related keywords<sup>13</sup> for our analysis consisting of 175 core covid related terms based on different criteria, e.g., health, policy, crisis, etc. Complete list can be found here<sup>13</sup>. In Fig. 1, we have also shown a small subset of our handpicked covid related features as example. For each keyword we have shown a category which intuitively represents a reasoning behind selecting this feature as relevant.

To verify if the chosen keywords are properly selected to represent covid we have conducted an LDA-based topic modeling analysis [2]. For the empirically selected topic number, LDA provides a word probability distribution for each topic from the corpus. So, each sorted word probability distribution set represents the most relevant terms to that topic from the corpus. For implementation, we have used python's **gensim** library. The algorithm to find the topics using LDA can be summarized as,

1. First, we made a document for each news article merging its headline, summary and news. As pre-processing, we converted it to lower cases, tokenized, lemmatized and removed stop words from it.
2. Then we fed each document to LDA and based on the empirically set parameters it returned topics.

We have set  $\alpha = 0.01, \eta = 0.01, pass = 4$  which are hyperparameters of the model and widely used. To find the suitable number of topics  $k$ , we conducted an exhaustive search by varying  $k$  between  $[2, 10]$  (a popular segment) and finding the  $k$  with highest  $c_v$  score or coherence score<sup>14</sup>. We found  $k = 10$  to give the highest  $c_v$  for both newspaper corpus (Guardian 0.55, Daily Star 0.52) and the topics were humanly differentiable. The top 30 words from the topic which represented covid in the Guardian and Daily Star are shown in Fig. 2 and 3 respectively. After finding the word distribution, we checked if our hand-selected core terms are properly representative of covid or not. We found that by looking at only the first 5000 words in the concerned extracted topic individually for both papers we could get 64% of our hand-selected terms in Guardian and 62% in Daily Star. So, the picked terms were quite relevant. Fig. ?? was generated using `pyLDAvis` library.

#### 4.2 Guardian (UK) and Daily Star (BD) discussed almost similar types of issues

To conduct this analysis for each newspaper, we merged all the articles into one large document. Then we removed the words which did not belong to our hand-picked features' list. So, finally We had a text document, which represented each features' appearance frequency in the collected news articles. Based on this, in Fig. 4, we have shown word cloud visualization for both newspapers. This visualization provides an interactive outlook to understand which terms have comparatively more occurrences. Using python's `WordCloud` library we exported the image. Seeing Fig. 4, we can get a grasp of the policies and center of focus for each country. From Fig. 4, we can see that both countries almost provided a similar type of focus on the concerned issues through more acknowledging them. Like here, we can see that both governments tried to impose **lockdown**, **travel ban**, focused on **social distancing**, **wearing masks**, etc as prevention mechanism. Lots of talks about **cases**, **deaths**, **unemployment**, **doctors**, etc. as different types of crisis. Talks about different types of **apps**, collaborative **research** as technological aspects to fight covid, etc. In Fig. 5, we have shown charts for some of the found highly frequent words for figurative understanding.

#### 4.3 Initial Highlight over Covid and Gradual Degrade in focus

Now, we will try to observe if there exists any sort of relationship between covid incident occurrences (cases and deaths variation) and the percentage of appeared covid related terms. The main intention was to see if the variation of covid incident creates an effect on the number of appeared covid related terms. For this, we imposed two different y-axes, the number of covid incident data and the percentage of appeared covid related terms for each week (x-axis), and shown the results in Fig. 6. The main reasoning behind seven days as the unit is, if we look at patterns in the covid incidents, we will observe that the variation

<sup>14</sup> <https://tinyurl.com/9f36682d>

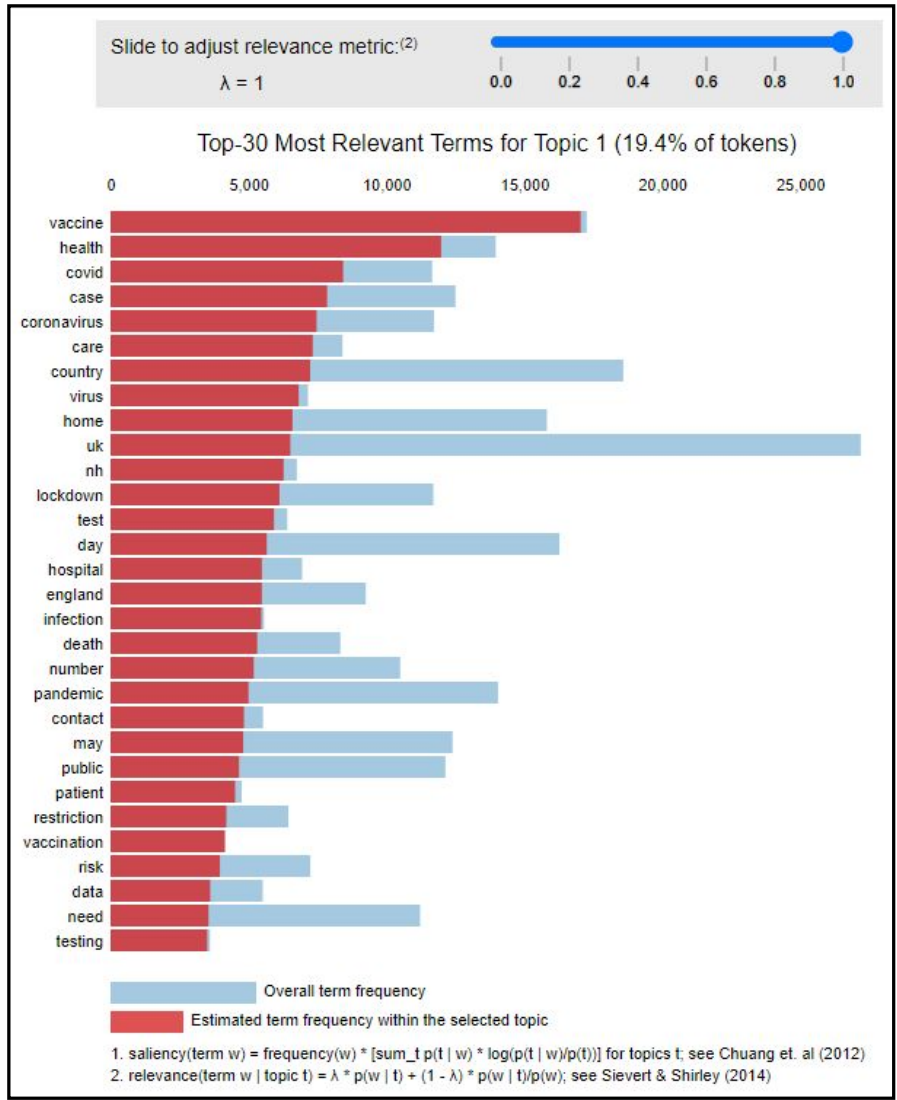


Fig. 2. First 30 words from the extracted topic as covid (Guardian)

is not abrupt rather slowly upwards or downwards [3]. So, this window might help to better understand associativity. From the corresponding figure, we get similar types of findings for both of the countries. Initially, a significant uprising in the covid related terms can be observed. Mostly due to the new appearance of an external event, new sets of strategies development, news to educate people about this new virus, good variation in covid incidents, etc. But gradually this



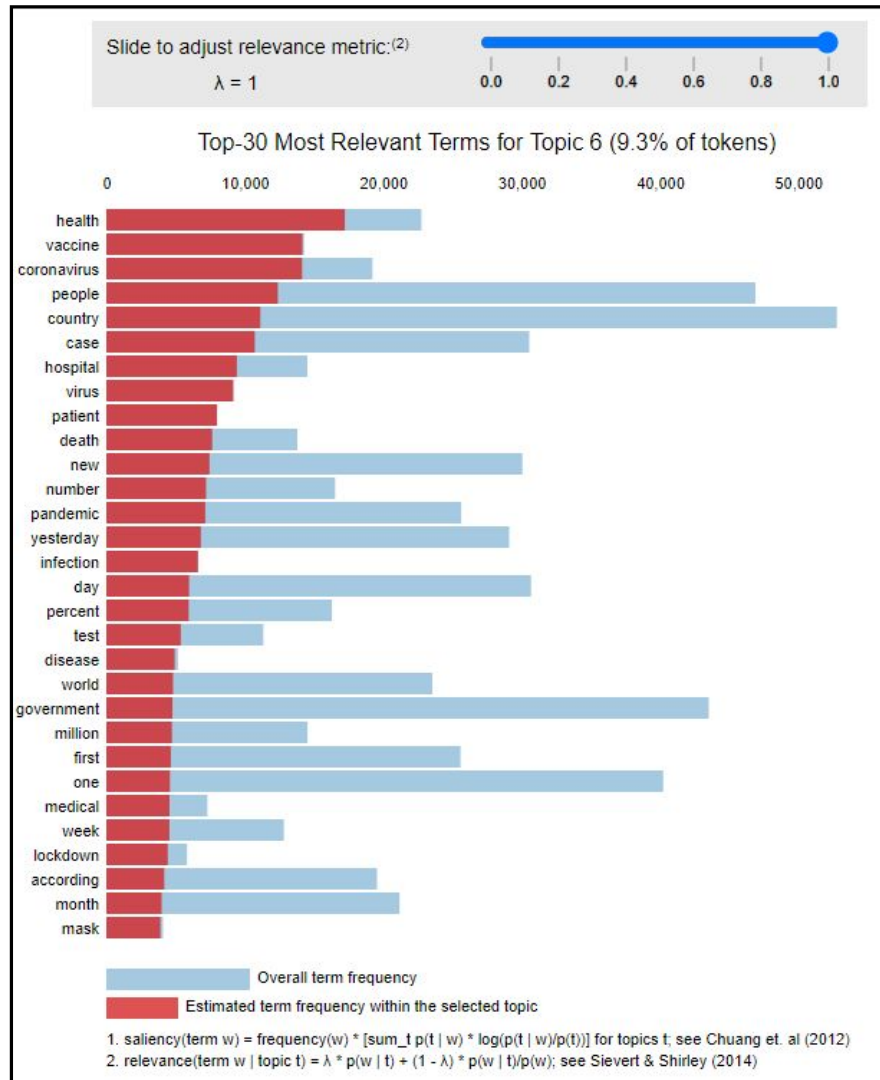


Fig. 3. First 30 words from the extracted topic as covid (Daily Star)

insurgence in covid terms gets reduced and takes a stable condition irrespective of good change in covid incidents. Mostly because it became a new normal event and people already knew a lot about it. This summarizes the portrayal of newspapers towards an external event.

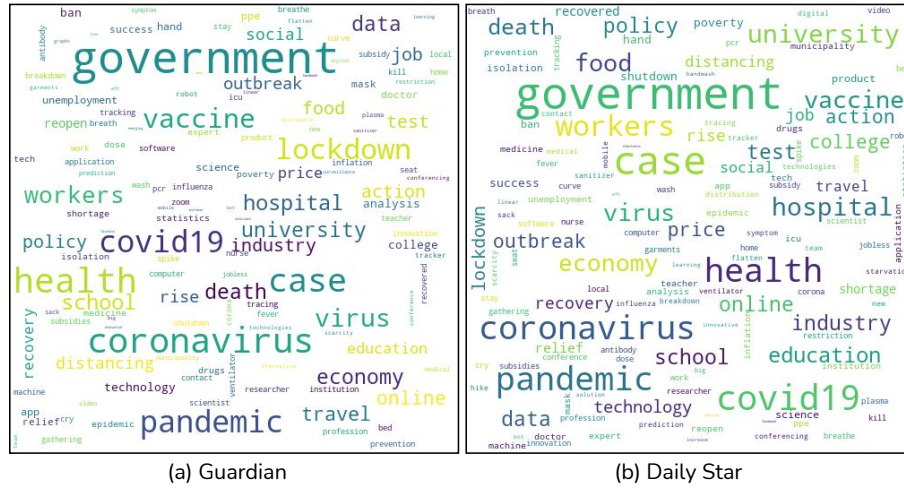


Fig. 4. Word cloud visualization

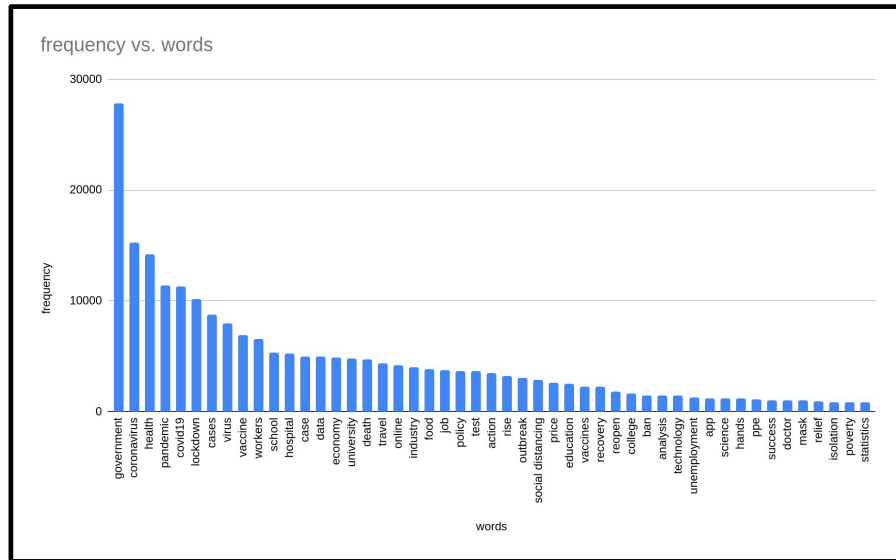
#### 4.4 Propensity of Warning vs comforting sentences

Warning (or negative) lines mostly represent different types of concerns, negative vibes and emotions, fearful situations, cursed words, etc and comforting (or positive) lines represent the opposite statements. The general characteristic of newspapers is to show more negative events compared to positive events [7] because these attract people more and help to construct social morals, to educate or aware people, etc. Here, our main goal was to observe, if the covid incident as an external event creates any special biasedness or effect over the ratio of warning and comforting sentences. Similar to previous charts, we show a week-by-week analysis with covid incidents in Fig. 7.

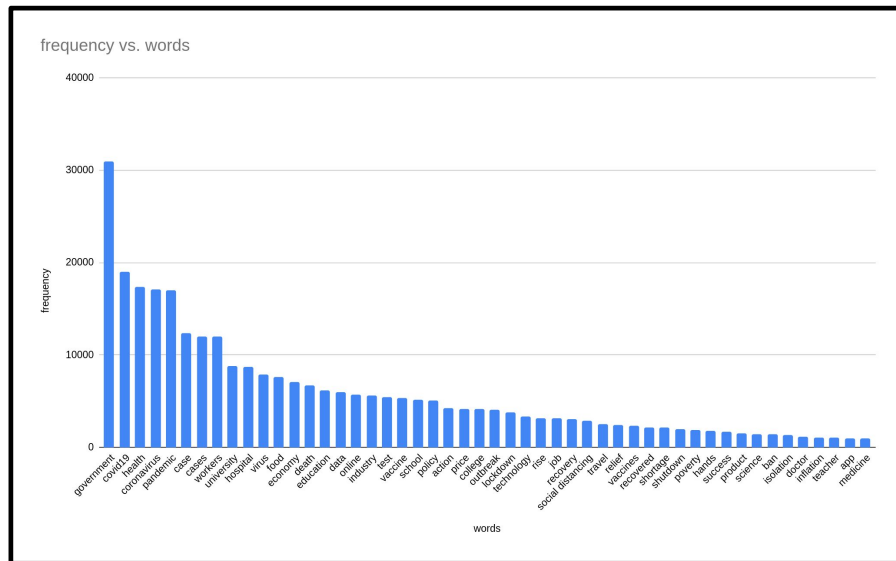
To determine the polarity of each appeared sentence in our data we have used Flair[1], a popular framework to do various NLP tasks. We have used their trained model for sentiment analysis which was trained over IMDB dataset <sup>15</sup>. In Fig. 8, we have shown how using Flair, we can get the polarity (POSITIVE or COMFORTING and NEGATIVE or WARNING) with corresponding confidence score for each sentence and using that information from each article we identify the number of comforting and warning sentences and finally perform a week-by-week analysis.

It was our initial expectation to see some insurgence in warning lines(%) after upward variation in covid incidents. If we look at the charts for the UK, we can see some small variation(around  $10^{th}$ - $15^{th}$ ,  $50^{th}$ - $60^{th}$  week) in warning lines(%) with the variation in recent(for that week) covid incidents but it is not too visible. But in the charts for BD, the visibility is quite vibrant and clear(around  $10^{th}$ - $30^{th}$ ,  $50^{th}$ - $60^{th}$  week). A possible reasoning can be if we observe the incident

<sup>15</sup> <https://ai.stanford.edu/~amaas/data/sentiment/>

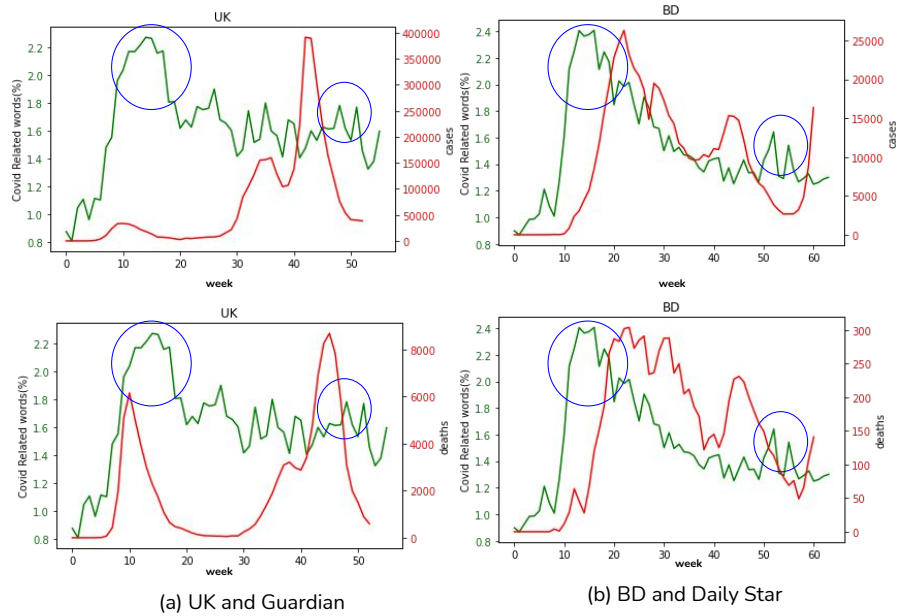


(a) Guardian



(b) Daily Star

Fig. 5. Frequency Distribution of Highly Frequent Top 50 Feature Terms



**Fig. 6.** Covid Incident vs Covid Related Terms(%)

charts of the UK, we can see that UK was comparatively more affected than Bangladesh with higher numbers. So, it is possible that throughout the year UK talked more about the negative aspects regarding covid, and thus when the insurgence in incidents occurred the ratio did not vary to a great extent. On the contrary, it is possible that Bangladesh talked more about the negative aspects when the crisis was more dominant. So, the variation was more vibrant.

## 5 Summary

Based on the three questions which we have addressed in section 1 and the discussion presented in section 4 we can summarize our findings as,

1. **How society gave importance over covid:** Both UK and BD, showed significant importance initially which reduced over time with it being a new normal event.
2. **Different types of portrayals of covid or which issues related to it were given more focus:** Both UK and BD, almost similarly addressed the relevant issues.
3. **How covid's severity controlled media's response (Warning(%) vs Comforting(%) Sentences:** The ratio of warning over comforting sentences was comparatively more dominant with the variation of covid incidents in the chart for BD than UK.

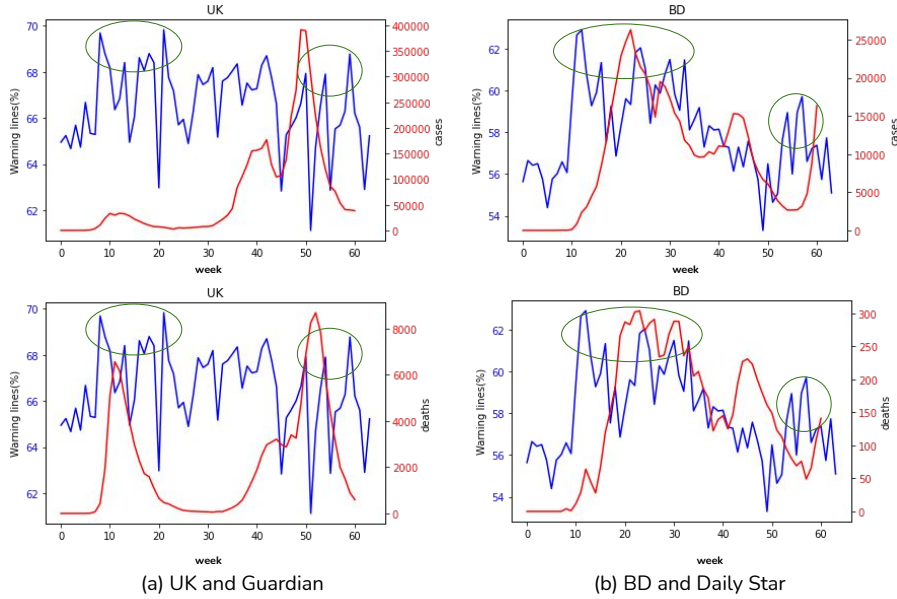


Fig. 7. Covid Incident vs Warning lines(%)

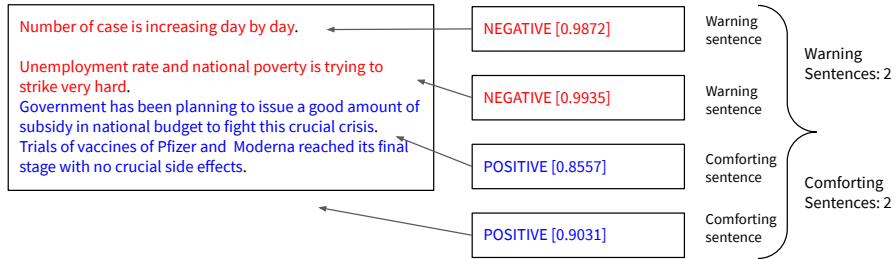


Fig. 8. Usage of Flair to label sentiment of each sentence

## 6 Conclusion

In this study, we have shown how newspapers react to an external event initially and its gradual degrading trend in focus based on an analysis of UK and BD’s covid situation and published news articles through applying different types of NLP tools and statistical measures. To summarize the intuition behind our findings, we can conclude as, the spread of covid and people’s awareness or relative response did not go exactly side by side. People’s actions are driven through different parameters and uncertainties. Covid created a significant amount of uncertainties and crisis in different aspects and layers of societies which have

driven people to perform different activities where were not exactly helpful to prevent the spread of covid. So, the main point which can be taken by policy-makers from this study is, the most important way to rise the awareness and support of people to fight covid is to reduce the uncertainties.

An important point to note is, the used mechanism to derive the findings is not exact. So, the findings can vary due to different types of data bias[4] and algorithm bias[6]. Also, to label the sentiments of sentences we have used a pre-trained model of Flair [1]. So, the biasedness occurred from the model is also included in the findings. In our work, we have hand-picked a set of feature terms that represented covid and used the pre-trained sentiment analysis model to conduct our experiments. As an ongoing work, we have plans to develop more compact strategies to discover a larger set of relevant terms and to design our sentiment labeling module to extract more in-depth knowledge related to covid from text data.

## References

1. Akbik, A., Blythe, D., Vollgraf, R.: Contextual string embeddings for sequence labeling. In: Proceedings of the 27th international conference on computational linguistics. pp. 1638–1649 (2018)
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. the Journal of machine Learning research **3**, 993–1022 (2003)
3. Dong, E.: Hongru du and lauren gardner. An Interactive Web-Based Dashboard to Track COVID-19 in Real Time pp. 30120–1
4. Greenland, S., Mansournia, M.A., Altman, D.G.: Sparse data bias: a problem hiding in plain sight. *bmj* **352** (2016)
5. Kang, B., Song, M., Jho, W.: A study on opinion mining of newspaper texts based on topic modeling. Journal of the korean society for library and information science **47**(4), 315–334 (2013)
6. Kuhlman, C., Jackson, L., Chunara, R.: No computation without representation: Avoiding data and algorithm biases through diversity. arXiv preprint arXiv:2002.11836 (2020)
7. Luo, J., Meier, S., Oberholzer-Gee, F.: No news is good news: CSR strategy and newspaper coverage of negative firm events. Harvard Business School (2011)
8. Rameshbhai, C.J., Paulose, J.: Opinion mining on newspaper headlines using svm and nlp. International Journal of Electrical and Computer Engineering (IJECE) **9**(3), 2152–2163 (2019)
9. Ta, T.H., Rahman, A.B.S., Sidorov, G., Gelbukh, A.: Mining hidden topics from newspaper quotations: The covid-19 pandemic. In: Mexican International Conference on Artificial Intelligence. pp. 51–64. Springer (2020)
10. Thirumaran, K., Mohammadi, Z., Pourabedin, Z., Azzali, S., Sim, K.: Covid-19 in singapore and new zealand: Newspaper portrayal, crisis management. Tourism Management Perspectives **38**, 100812 (2021)