

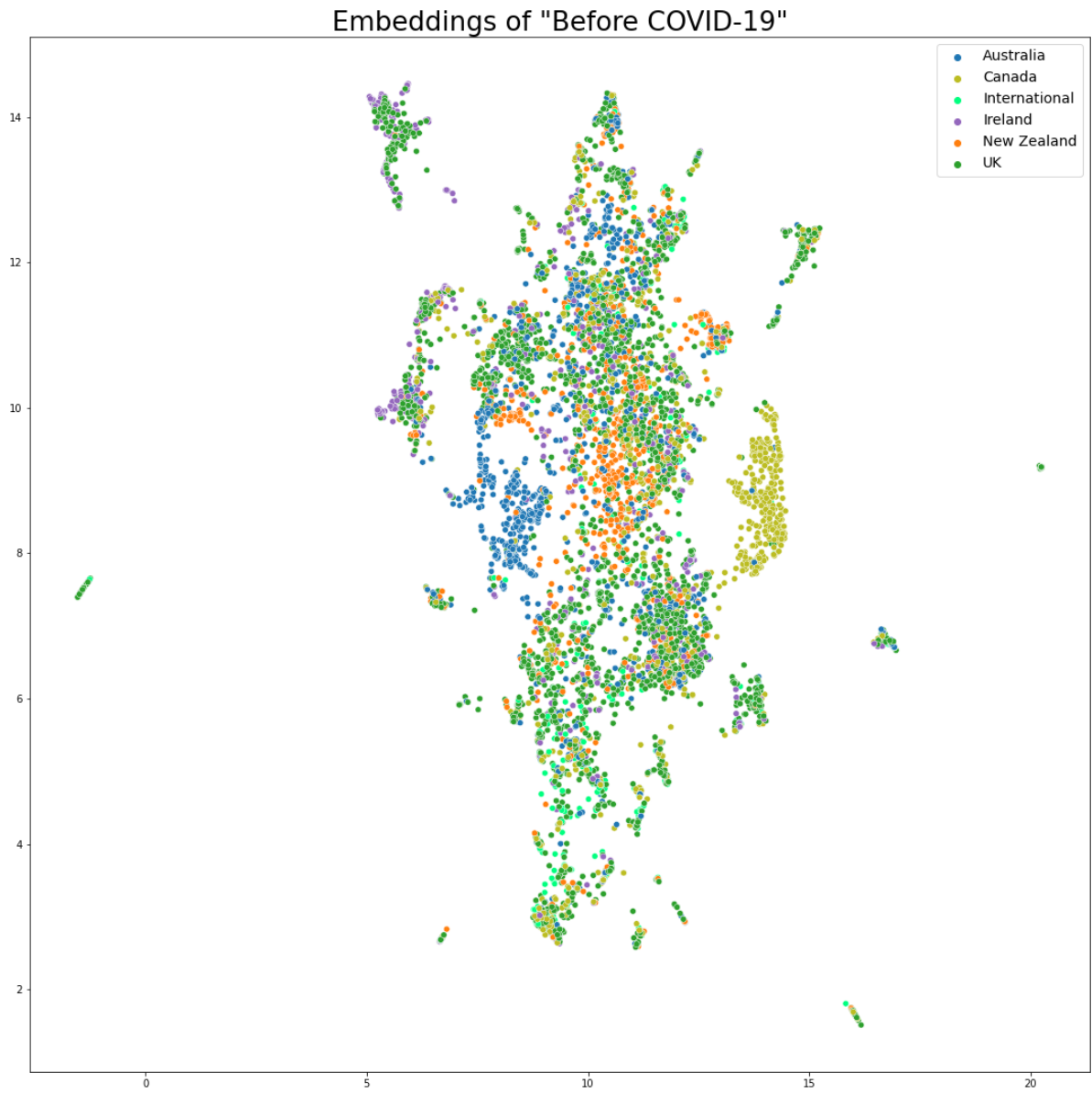
Supplementary

1. Topic Detection

Choosing the optimal model

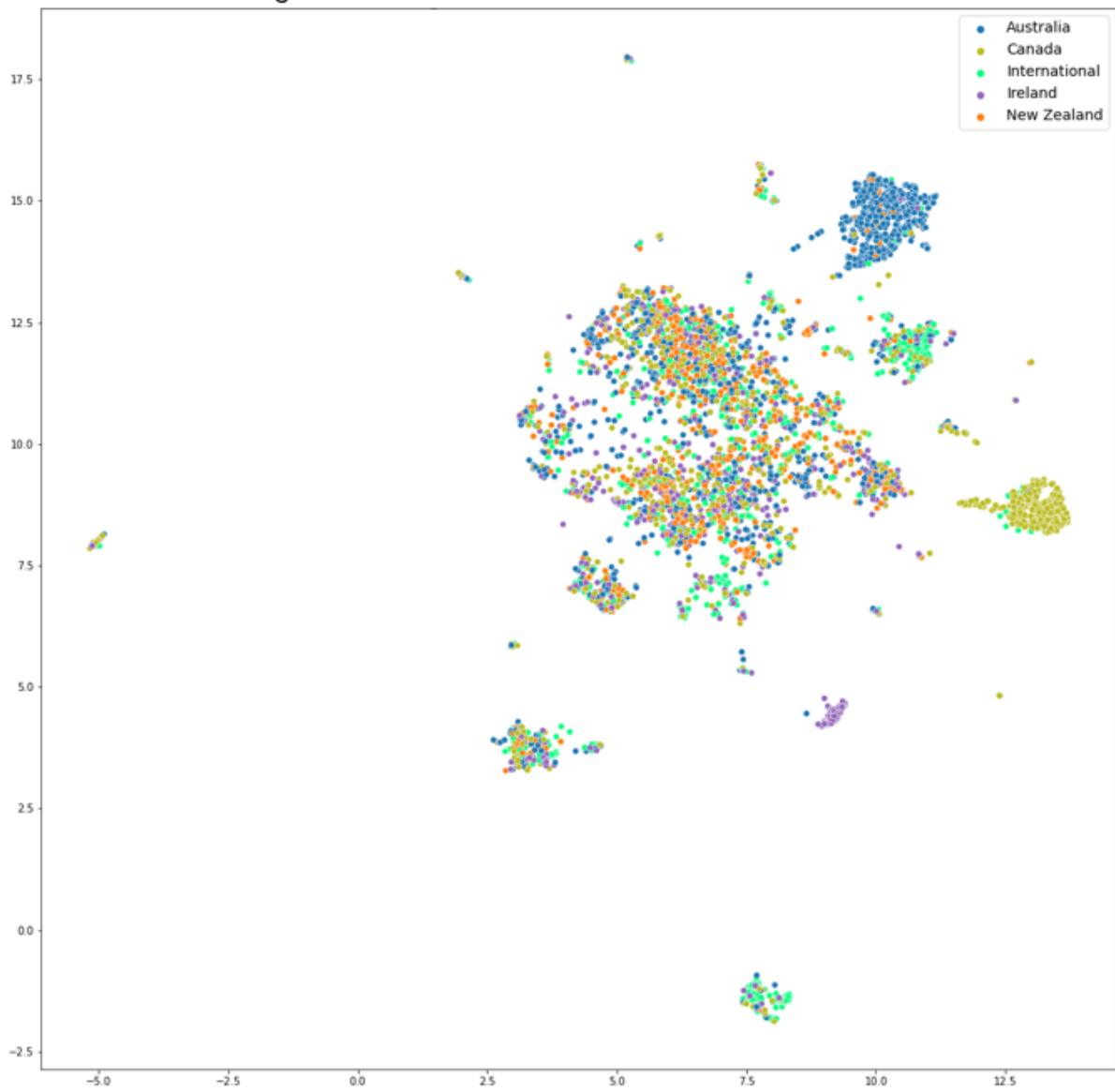
The scatter plots in figures 1, 2 and 3 show how the embeddings of the headlines are scattered with respect to each other for the three different periods. The scatter in each plot are colour coded, with respect to which country they are from. Figure 1, 2 and 3 clearly show, there are topics only dealing with Australia and Canada with respect to vaccines in all three periods. In “After the COVID-19 vaccine announcement” there is also a small cluster containing headlines about Ireland and vaccines. Particularly in “Before COVID-19” there does only seem to be a couple of clear clusters of headlines in the data, while there seem to be areas with more dense scatter than other areas.

Figure 2 indicates that the period “Before the COVID-19 vaccine announcement” has perhaps five or seven clear clusters depending on the minimum size of a cluster. The big multicoloured area in the middle is though characterised by being more dense in some areas, indicating that it could be a collection of several topics. The structure of figure 3 is somewhat similar to figure 2, it has five clusters of headlines, which stands out from the remaining and a multicoloured group of headlines with a varied density within.

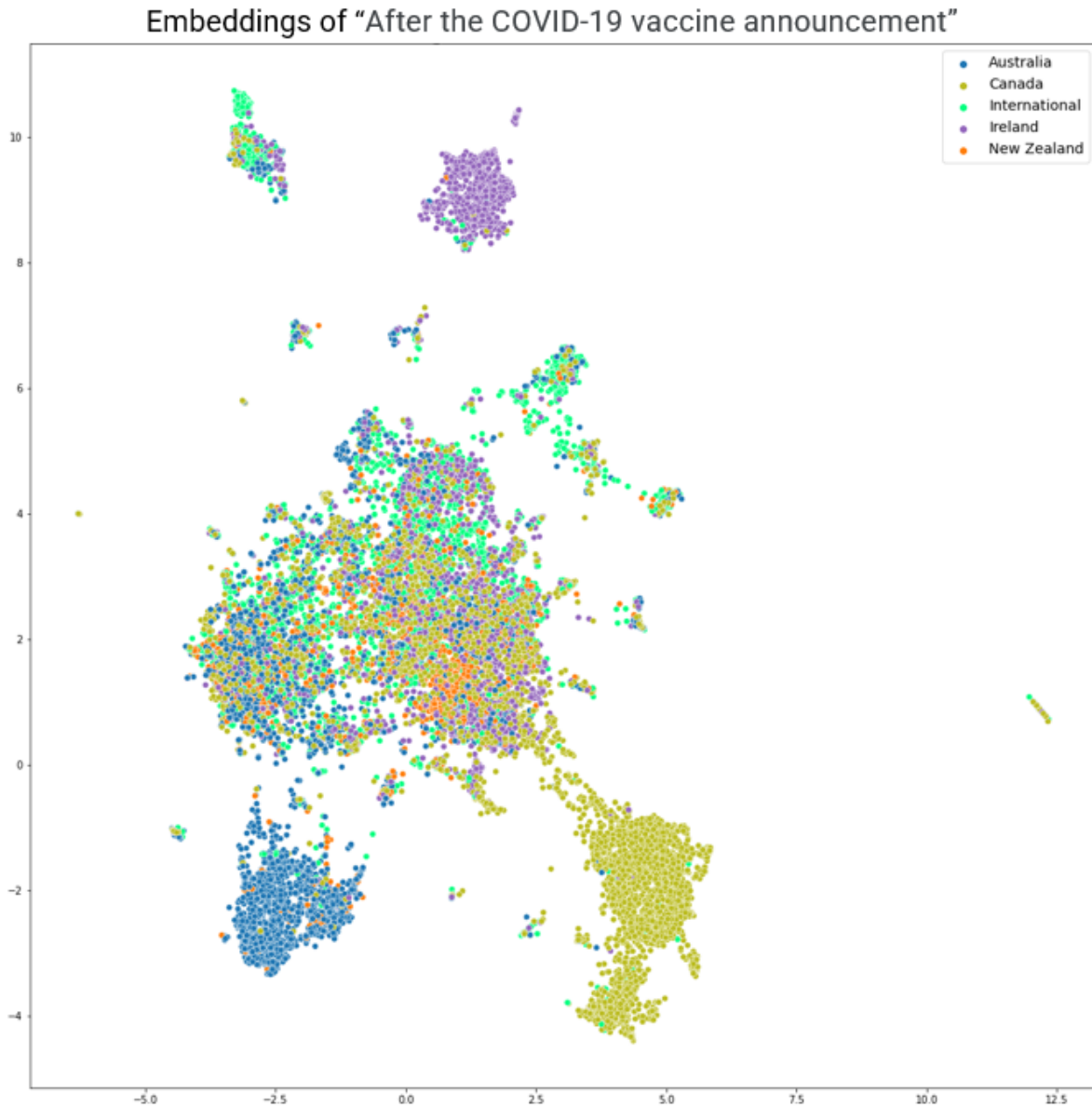


Supplementary Figure 1: Embedding for before COVID-19

Embeddings of "Before the COVID-19 vaccine announcement"



Supplementary Figure 2: Embedding for "Before the COVID-19 vaccine announcement"



Supplementary Figure 3: Embedding for "After the COVID-19 vaccine announcement".

2. Analysis of n-grams

Method

Topics for the different vaccines or vaccine manufacturers

The results from NER were used to find the most frequently occurring COVID-19 vaccine manufacturers or specific vaccines. The seven most frequently used by both English and

non-English ONSs were: “AstraZeneca”, “BioNTech”, “Johnson & Johnson”, “Moderna”, “Oxford”, “Pfizer” and “Sputnik V”. These will for simplicity be referred to as vaccines in the remainder of the report.

To learn if a possible change in sentiment from “Before the COVID-19 vaccine announcement” to “After the COVID-19 vaccine announcement” with respect to the individual vaccines was caused by certain developments or changes over time, an assessment of the different vaccines were done for both periods. For this the lemmatized headlines created for the topic detection were stemmed and used. This made sure that only words with some deeper meaning were used and that the same words but with different endings were assessed as the same.

The assessment was initially done counting frequent words, but it was difficult to derive anything useful from these. Based on this frequent bigrams and trigrams were counted instead. These were derived from the headlines using “bigrams” and “trigrams” from Natural Language Toolkit (NLTK). These will subsequently have the common name n-grams.

As the datasets contained different amounts of headlines, it was difficult to compare the individual frequent bigrams and trigrams the vaccines and periods between. Based on this, the n-grams remaining were used to count the number of headlines each n-grams was contained in, which was then used to calculate the given n-grams relative frequency with respect to the number of headlines in the dataset using equation 5.1.

The 30 most frequent n-grams for each dataset were used to make a barplot, illustrating the percent of headlines each n-gram was contained in. This was done using “barplot” from “seaborn”.

Results

The creation of subsets for all vaccines resulted in a total of 21 datasets, where each contained the data from “Before COVID-19”, “Before the COVID-19 vaccine announcement” or “After the COVID-19 vaccine announcement”.

This shows a change in media coverage with respect to each period and vaccine. Sputnik V is the name of a COVID-19 vaccine with no mention of it before the COVID-19 outbreak. The remaining vaccines are actually companies developing a COVID-19 vaccine, but there has definitely also been an increase in media attention towards these the periods between. The three periods do not have the same length, which does not matter, as the longest period is

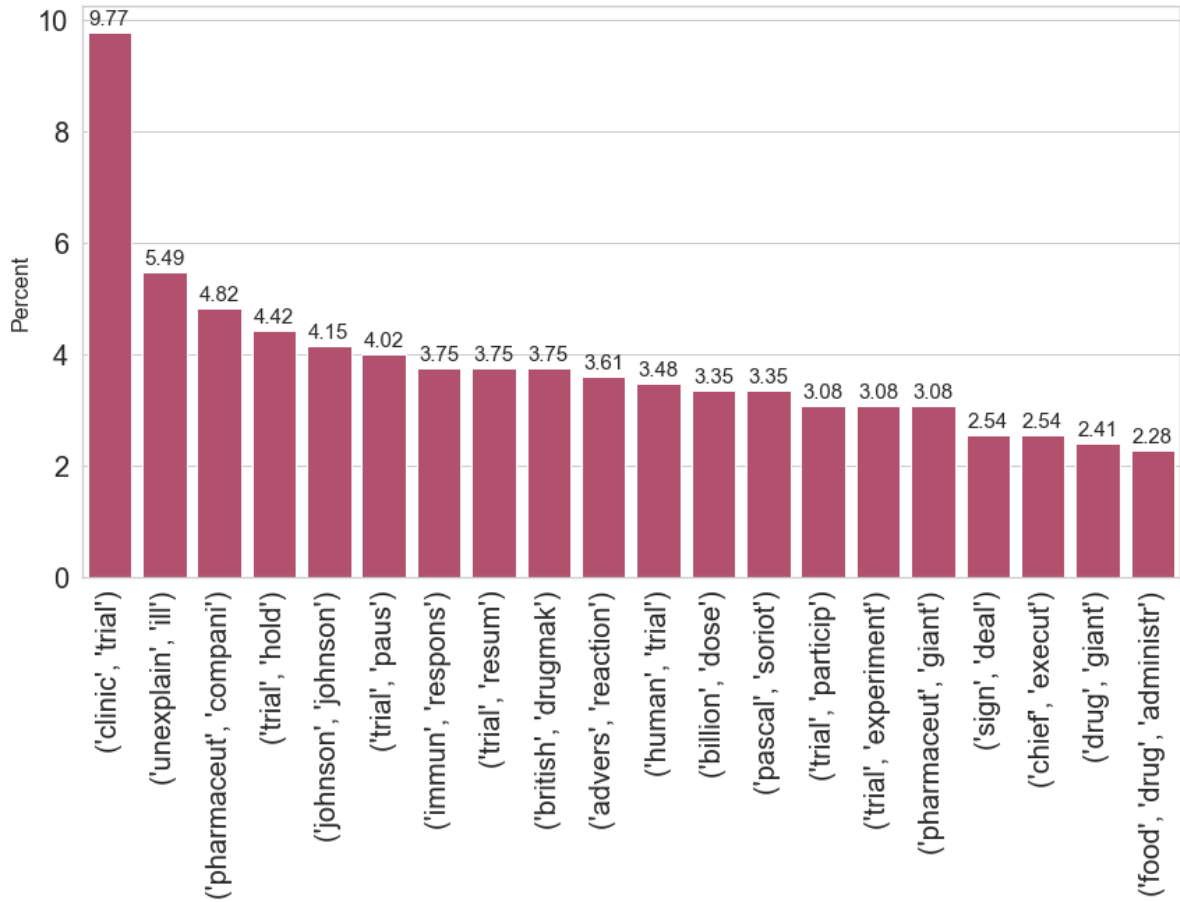
“Before the COVID-19 vaccine announcement”, which has the fewest observations, while the shortest is “After the COVID-19 vaccine announcement”, which has the most.

The most frequent n-grams, with respect to the individual vaccines, show some similarities while also showing some clear differences. The results for companies, which have developed their vaccine jointly, are so similar, that they will be assessed together to not repeat the same things several times, while the remaining companies will be assessed individually. Common for all vaccines is that they in “Before the COVID-19 vaccine announcement” all bear a clear sign of being in a trial phase, as they all have several n-grams within the 20 most frequent containing words like: “trial” and “test”. In “After the COVID-19 vaccine announcement” all mention the “European Medicines Agency” within their 20 most frequent n-grams, while except for Sputnik V, the remaining also mention South Africa.

AstraZeneca and Oxford As AstraZeneca and Oxford have developed their vaccine jointly, it makes sense, if they have been mentioned in several headlines together. A count showed that they had 1901 headlines in common, whereby multiple n-grams had been counted with respect to both of them. In figures 4 and 6 it is clear, that for “Before the COVID-19 vaccine announcement” AstraZeneca’s and Oxford’s most frequent n-grams are heavily influenced by the fact, that they were in a trial phase, this is among other seen in the most frequent n-gram for both of them, which occur in 8.71 percent of Oxford’s headlines and 9.77 percent of AstraZeneca’s. The period though also carry sign of problems with the vaccine during the trial phase, as both of them have the frequent bigrams “(trial, hold)” and “(trial, paus)”, while AstraZeneca also have “(trial, resum)”.

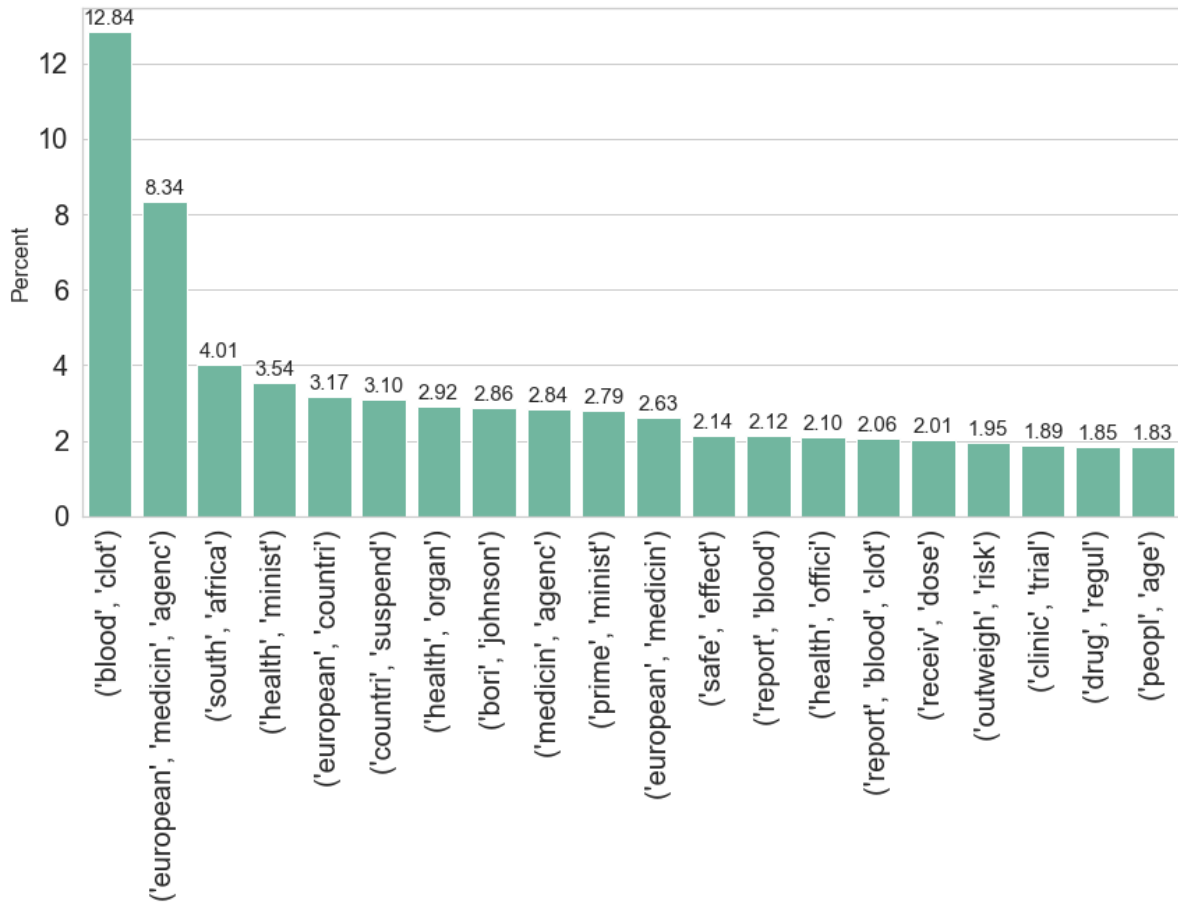
The frequent n-grams for “After the COVID-19 vaccine announcement” show that both companies experience some media attention with respect to blood clots. While blood clots are mentioned in 12.84 percent of the headlines also mentioning AstraZeneca, it is only in 5.65 percent of headlines with respect to Oxford. Which is almost 5 times as many headlines mentioning the combination AstraZeneca and blood clots than the combination Oxford and blood clots.

Frequent n-grams AstraZeneca
 "Before the COVID-19 vaccine announcement"



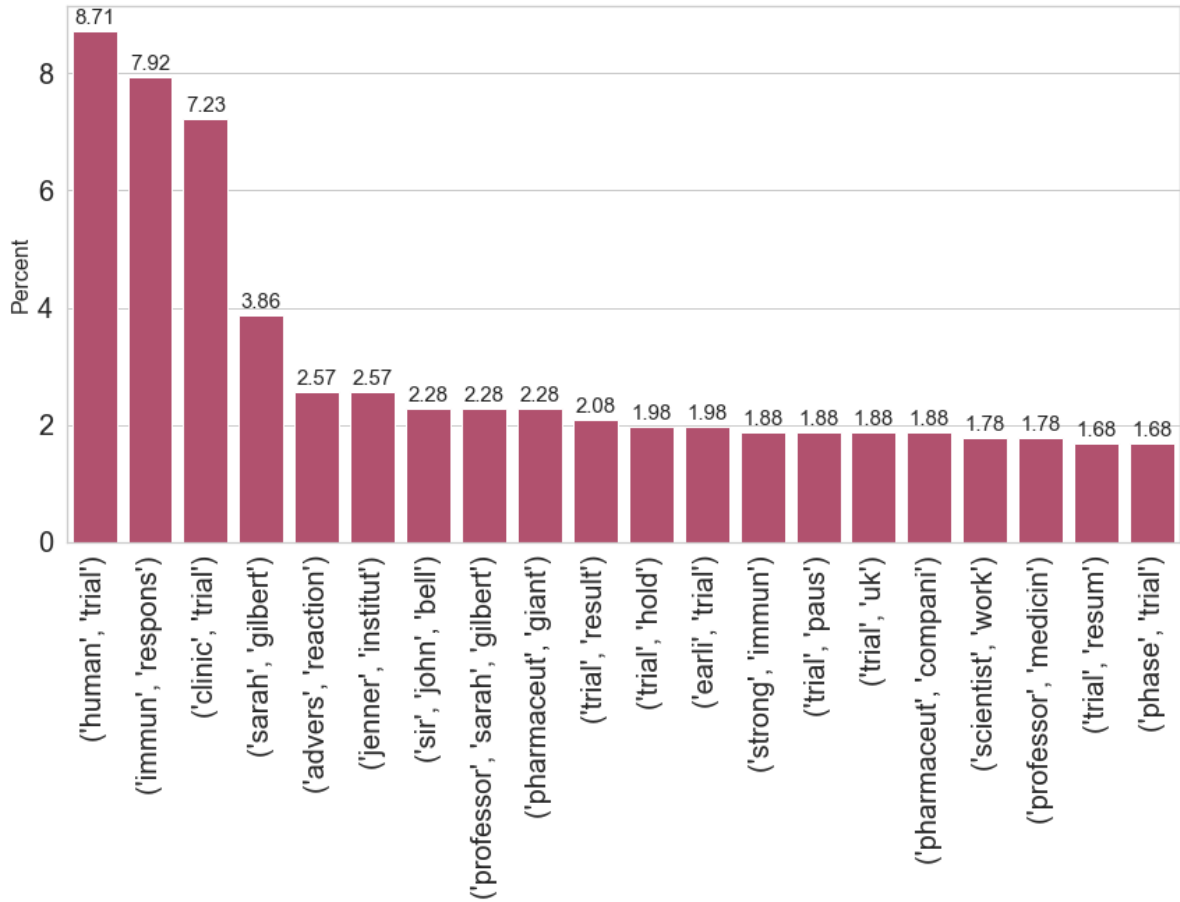
Supplementary Figure 4: The 20 most frequent bigrams and trigrams in headlines which are about AstraZeneca for the period "Before the COVID-19 vaccine announcement".

Frequent n-grams AstraZeneca
 "After the COVID-19 vaccine announcement"

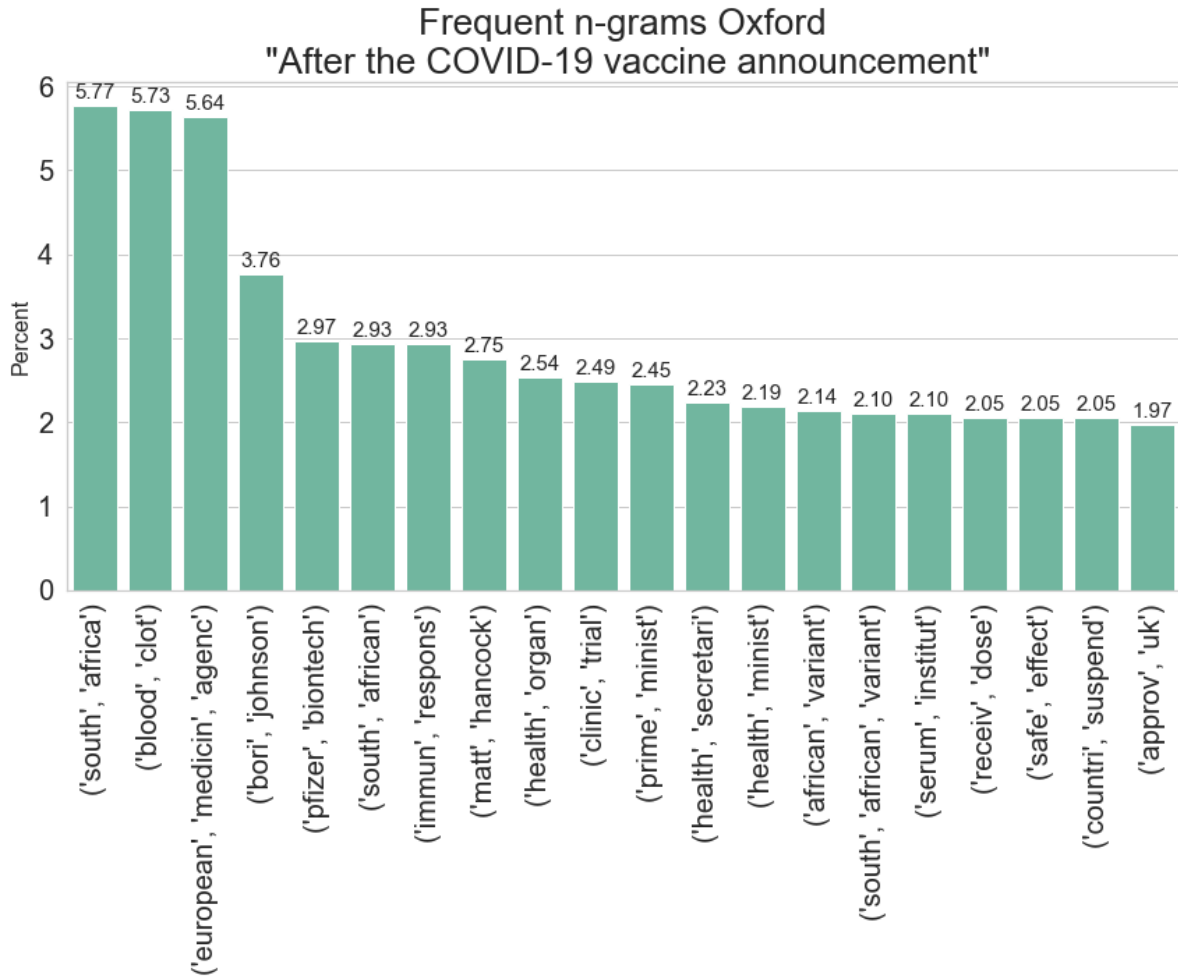


Supplementary Figure 5: The 20 most frequent bigrams and trigrams in headlines which are about AstraZeneca for the period "After the COVID-19 vaccine announcement".

Frequent n-grams Oxford
 "Before the COVID-19 vaccine announcement"



Supplementary Figure 6: The 20 most frequent bigrams and trigrams in headlines which are about Oxford for the period "Before the COVID-19 vaccine announcement".

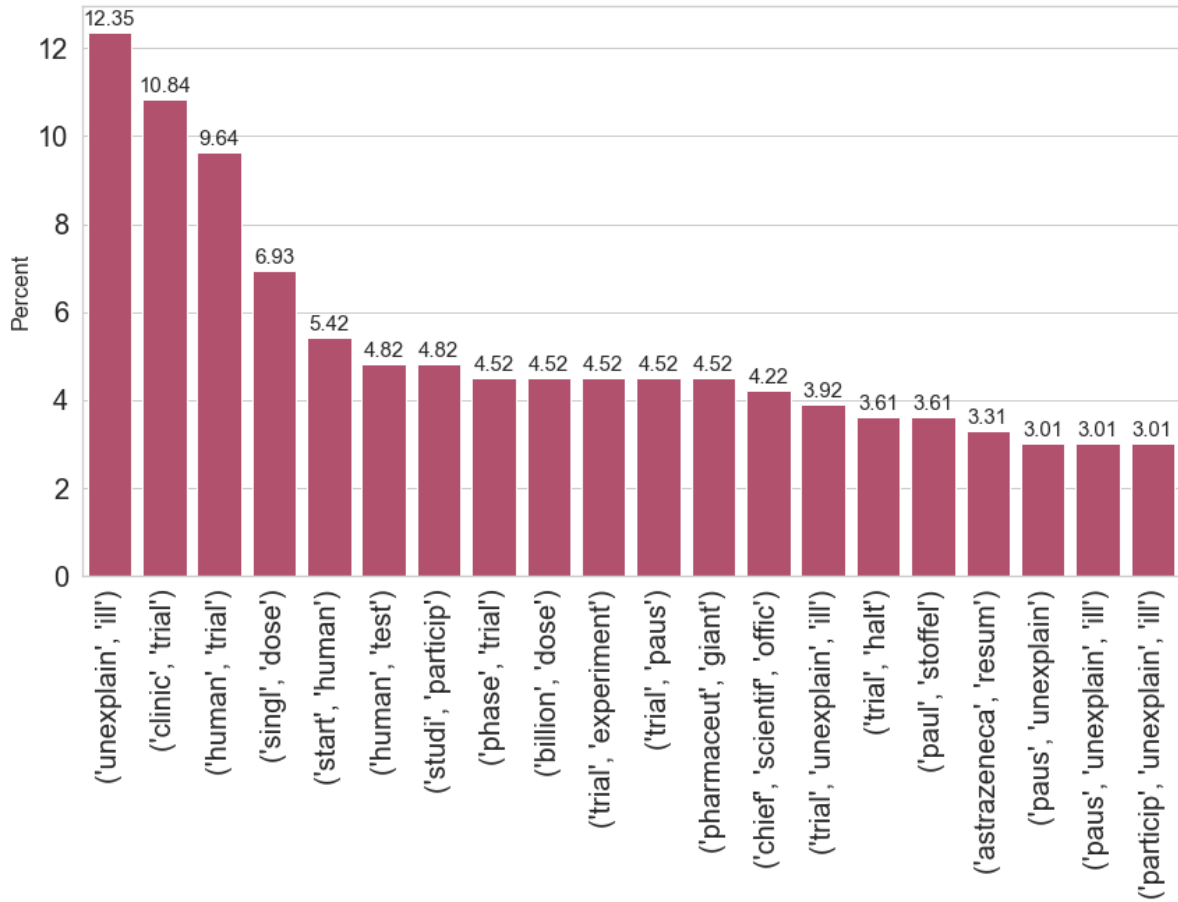


Supplementary Figure 7: The 20 most frequent bigrams and trigrams in headlines which are about Oxford for the period "After the COVID-19 vaccine announcement".

Johnson & Johnson Figures 8 and 9 illustrate the frequent n-grams with respect to Johnson & Johnson. For "Before the COVID-19 vaccine announcement" there seem to be two tendencies. The first is that Johnson & Johnson are in a trial phase like all other vaccines. Additionally the period seems to carry a lot of weight with respect to unexplained illness, which caused their trial to pause. This can amongst others be seen in the bigrams: "(unexplained, ill)", "(trial, paus)", "(trial, halt)" etc. The unexplained illness in their study have had a lot of media attention as the bigram "(unexplain, ill)" occurs in 12.35 percent of the Johnson & Johnson headlines for this period.

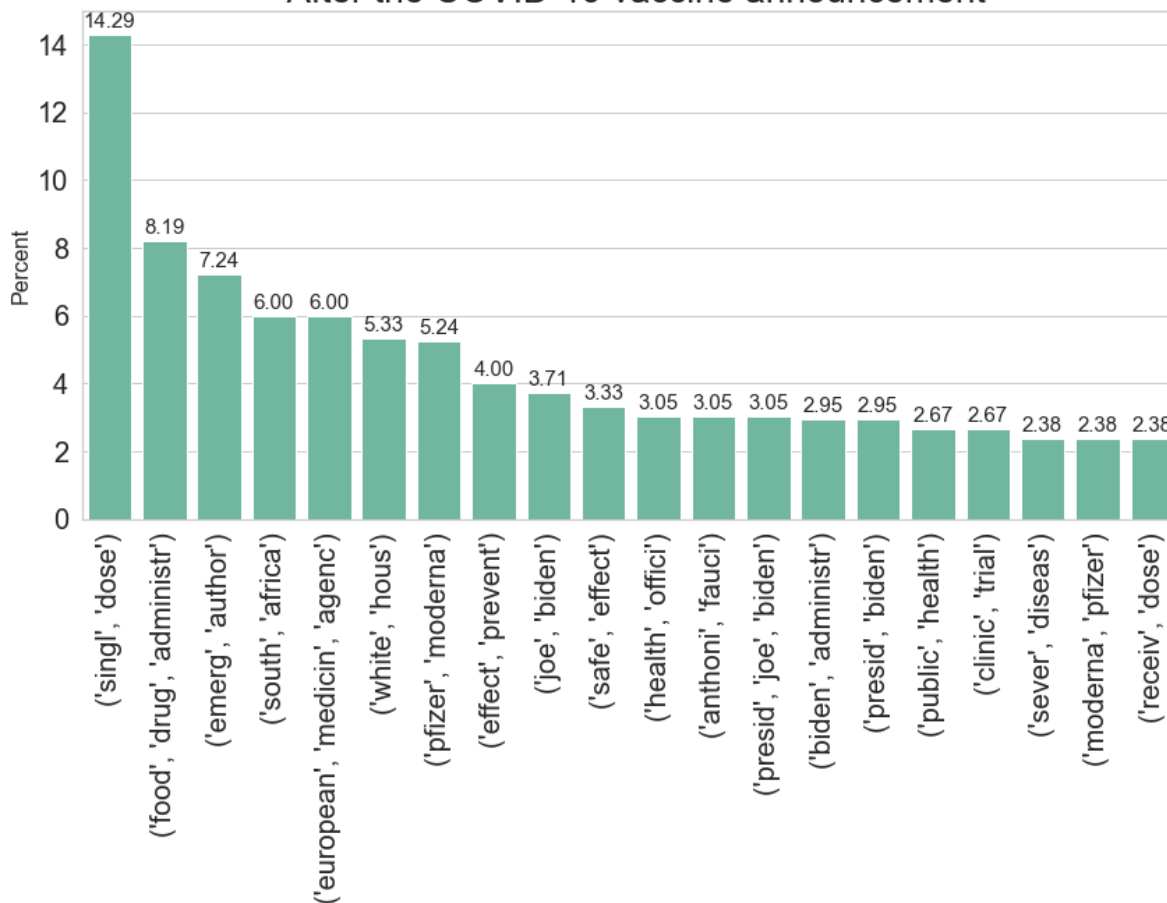
In "After the COVID-19 vaccine announcement" the media have a lot of focus on the fact, that Johnson & Johnson in contrast to any of the other already approved vaccines is a single dose vaccine, which is mentioned in 14.90 percent of the headlines.

Frequent n-grams Johnson & Johnson
 "Before the COVID-19 vaccine announcement"



Supplementary Figure 8: The 20 most frequent bigrams and trigrams in headlines which are about Johnson & Johnson for the period "Before the COVID-19 vaccine announcement".

Frequent n-grams Johnson & Johnson "After the COVID-19 vaccine announcement"



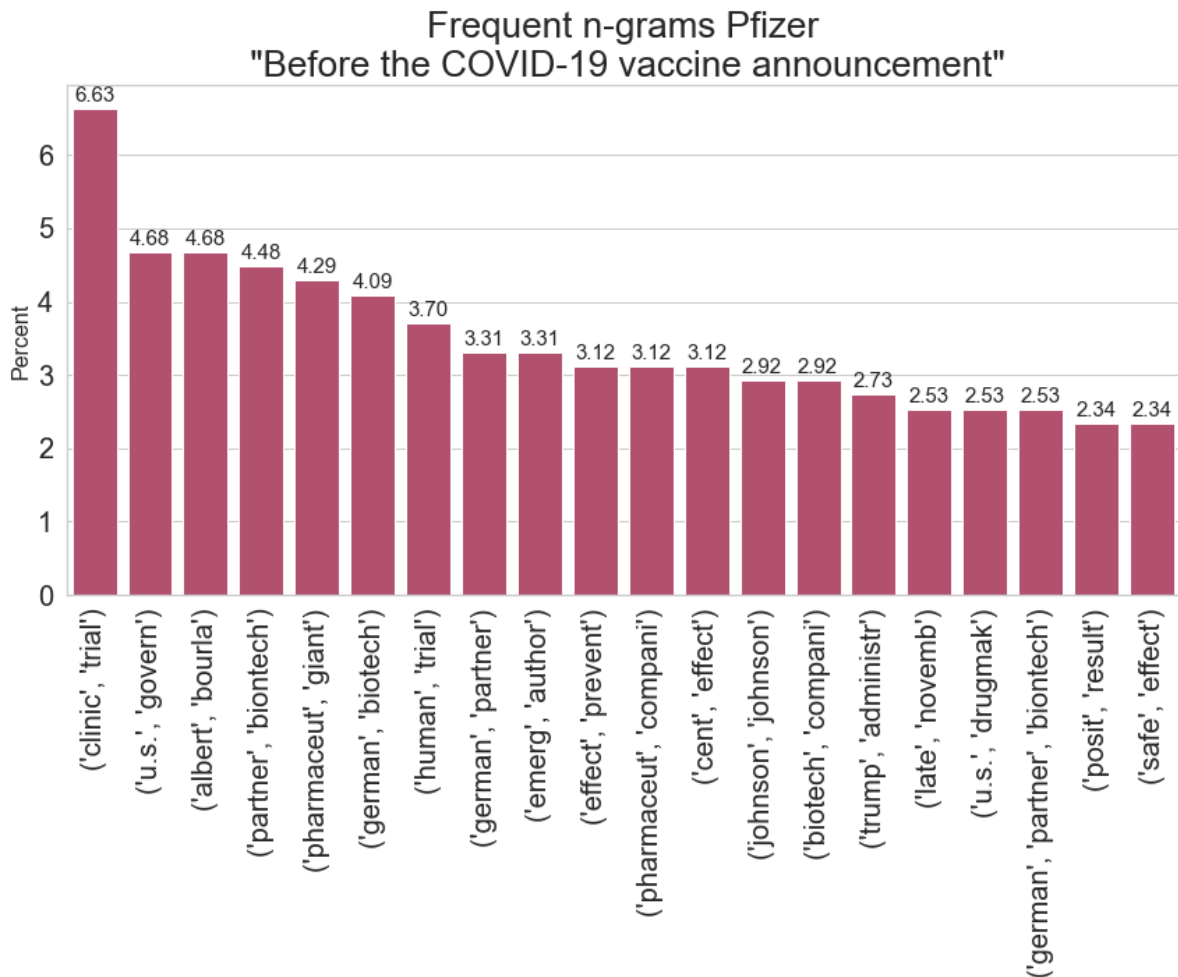
Supplementary Figure 9: The 20 most frequent bigrams and trigrams in headlines which are about Johnson & Johnson for the period "After the COVID-19 vaccine announcement".

Pfizer and BioNTech like AstraZeneca and Oxford, Pfizer and BioNTech have developed their vaccine jointly, whereby it likewise makes sense, that they have a lot of frequent n-grams with respect to their headlines in common. A search showed that 2,125 headlines did actually mention them both.

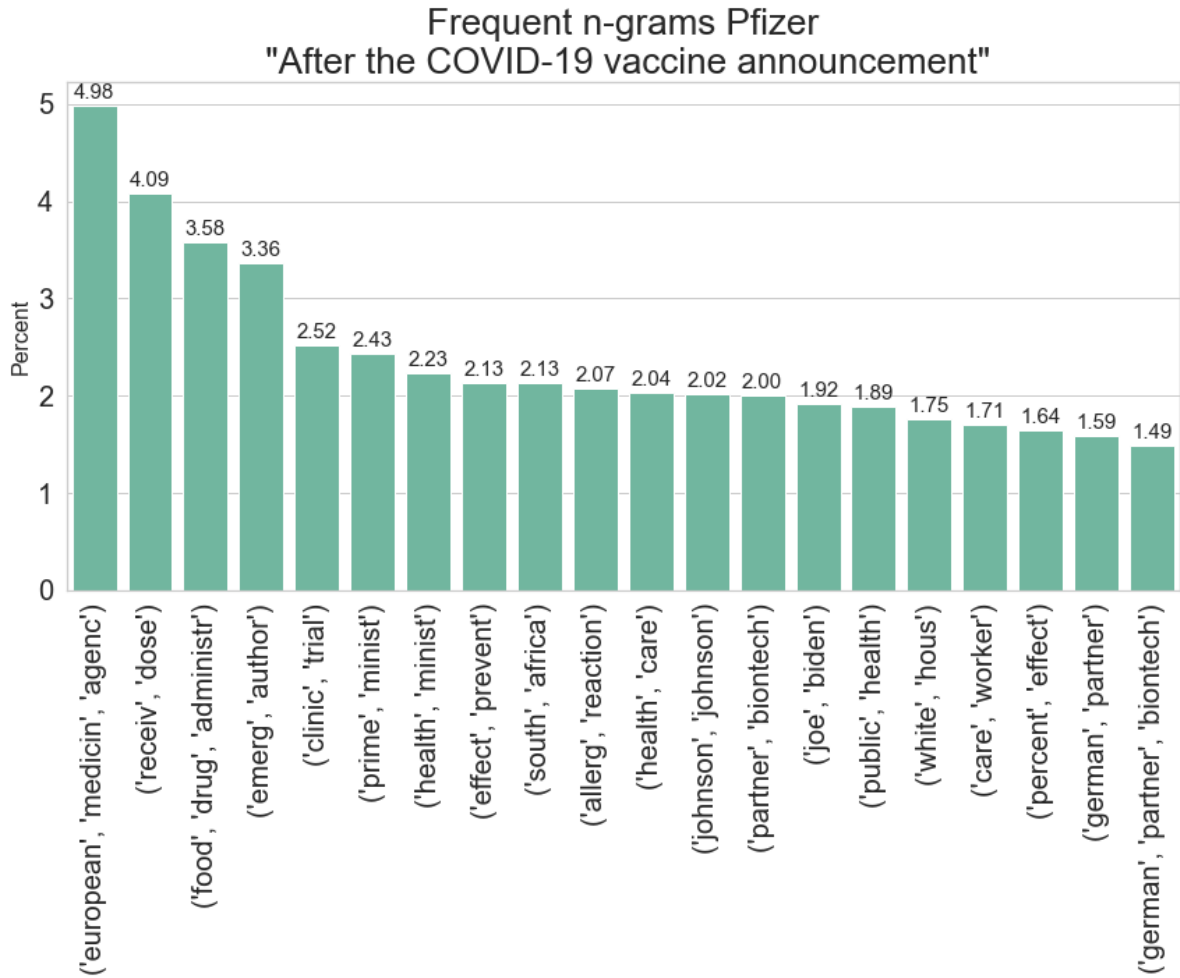
Like all other vaccines "Before the COVID-19 vaccine announcement" is characterised by being a trial period. Additionally it show signs of an upcoming approval, with n-grams like "(fast, track)", "(fast, track, status)", "(sign, deal)" and "(regulatory, approval)" being frequent n-grams in BioNTech, while "(emerg, author)" and "(late, novemb)" are indicators for Pfizer, this can be seen in figures 10 and 12.

In "After the COVID-19 vaccine announcement" Pfizer and BioNTech both have relative low frequency with respect to their most frequent n-grams, where the two most frequent for both of them are "(european, medicin, agenc)" and "(receiv, dose)", which for Pfizer occur in 4.94

percent and 4.08 percent of the headlines respectively, while in BioNTech's headlines they occur in 7.11 percent and 4.82 percent respectively. In this period there does not seem to be any particular subject, which is in focus, with respect to the media for any of the two vaccines.

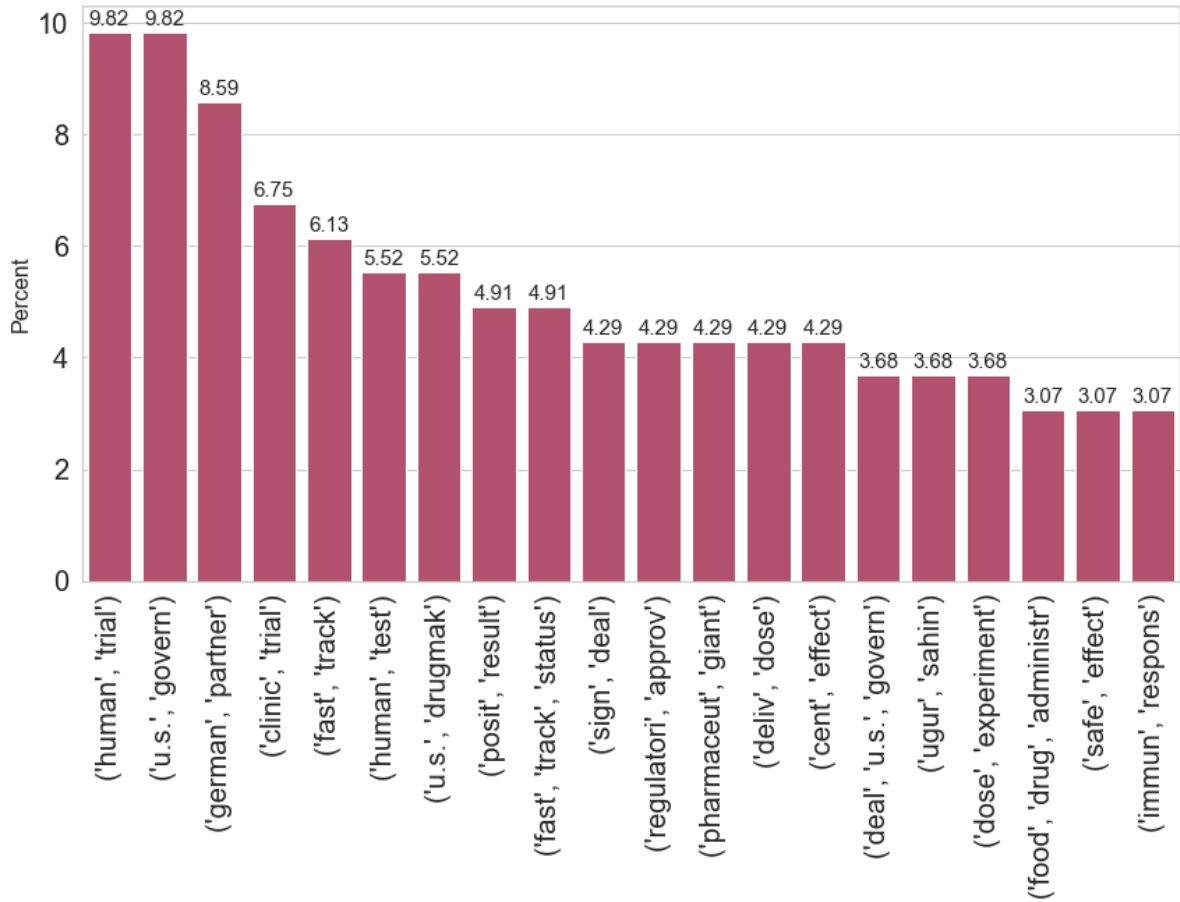


Supplementary Figure 10: The 20 most frequent bigrams and trigrams in headlines which are about Pfizer for the period "Before the COVID-19 vaccine announcement".

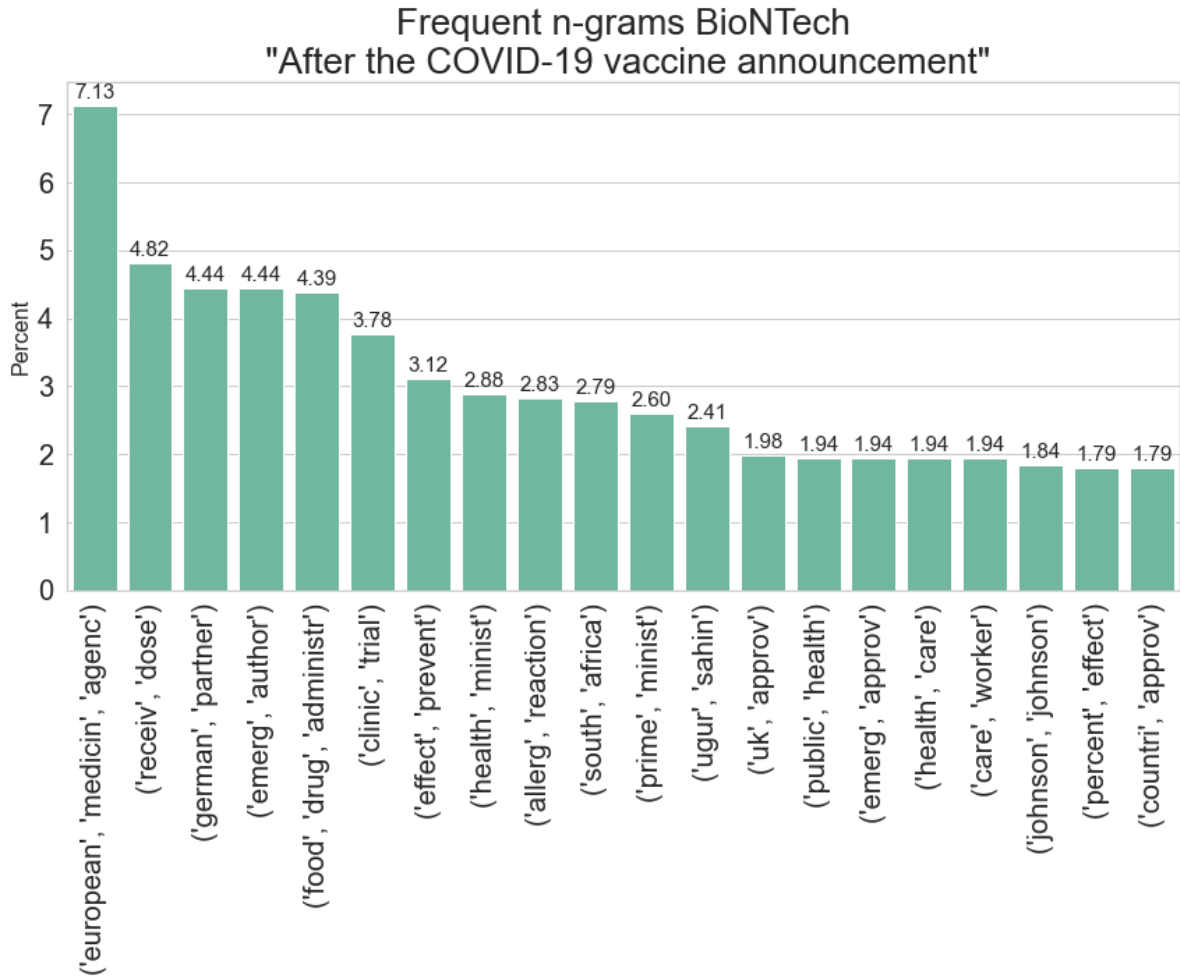


Supplementary Figure 11: The 20 most frequent bigrams and trigrams in headlines which are about Pfizer for the period "After the COVID-19 vaccine announcement".

Frequent n-grams BioNTech
 "Before the COVID-19 vaccine announcement"



Supplementary Figure 12: The 20 most frequent bigrams and trigrams in headlines which are about BioNTech for the period "Before the COVID-19 vaccine announcement".

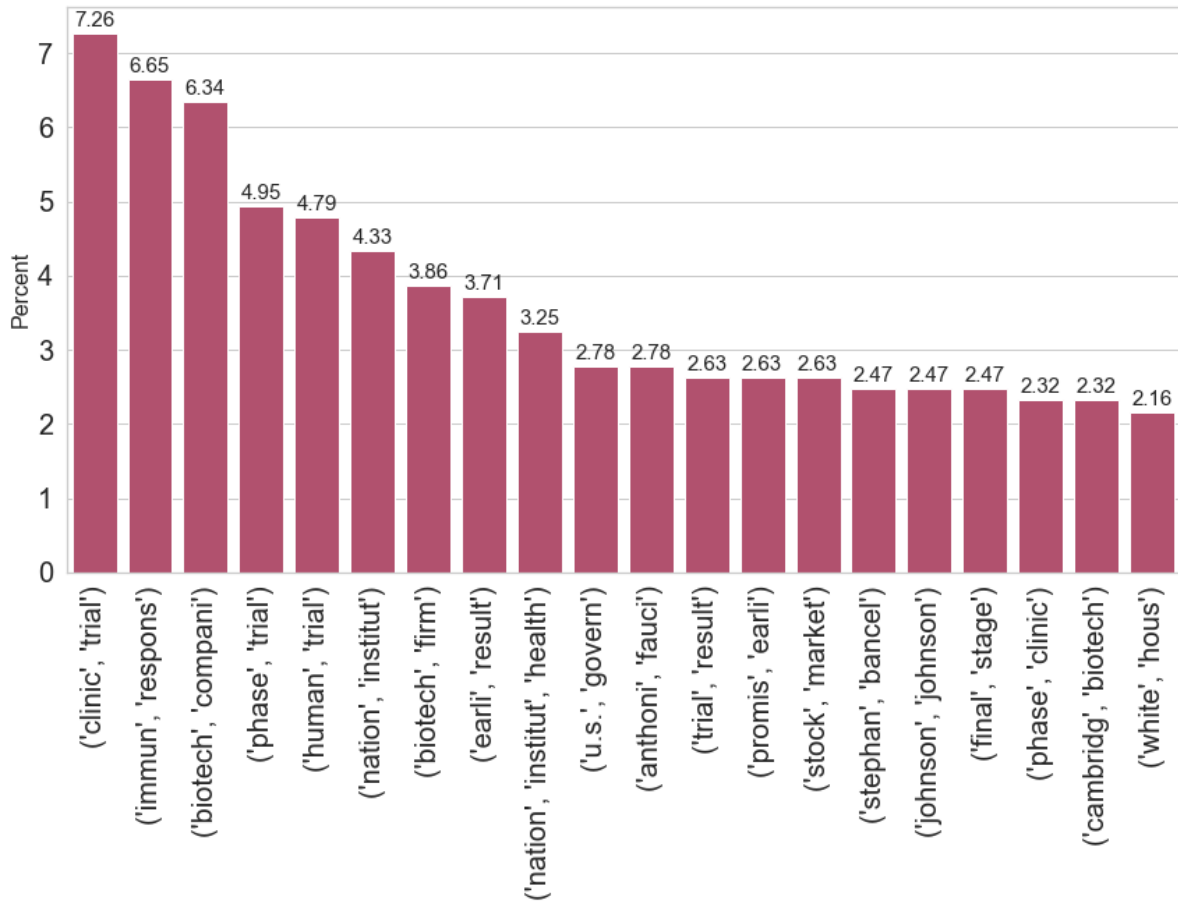


Supplementary Figure 13: The 20 most frequent bigrams and trigrams in headlines which are about BioNTech for the period "After the COVID-19 vaccine announcement".

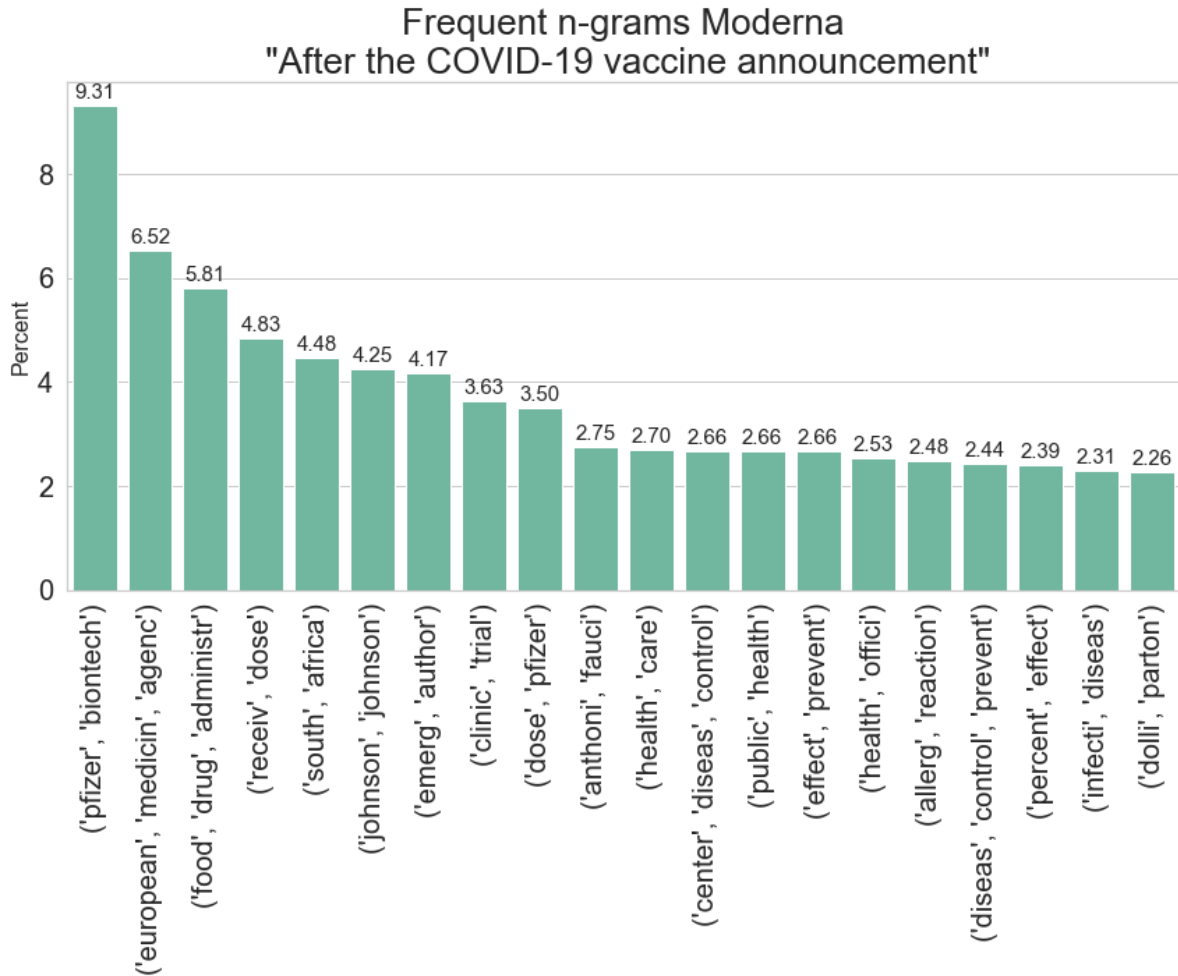
Moderna In "Before the COVID-19 vaccine announcement" Moderna does not show any sign of other subjects getting some high media attention with respect to the frequent n-grams, other than being in a trial phase, this can be seen in the 20 most frequent n-grams, illustrated in figure 14. This also shows that in 3.25 percent of the headlines, Moderna is mentioned in combination with "(nation, institut, health)", whereby it is the only vaccine having NIH among its most frequent n-grams.

For "After the COVID-19 vaccine announcement" the bigram "(pfizer, biontech)" is the most frequent occurring in 9.29 percent of the headlines. Additionally it is the only vaccine having "(center, diseas, control)" among its 20 most frequent n-grams. 2.27 percent of the headlines in this period also mention Dolly Parton, which probably is caused by her being one of their major funders.

Frequent n-grams Moderna
 "Before the COVID-19 vaccine announcement"



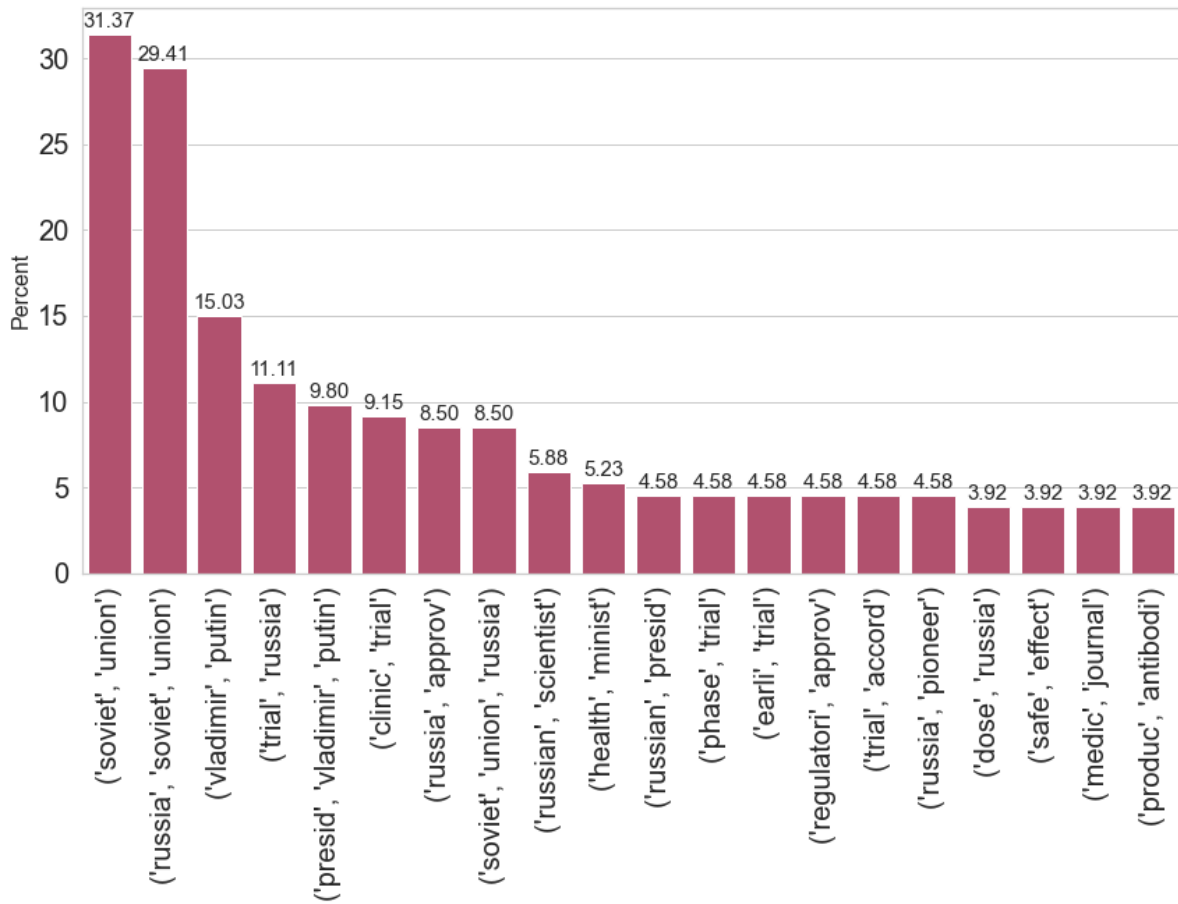
Supplementary Figure 14: The 20 most frequent bigrams and trigrams in headlines which are about Moderna for the period "Before the COVID-19 vaccine announcement".



Supplementary Figure 15: The 20 most frequent bigrams and trigrams in headlines which are about Moderna for the period "After the COVID-19 vaccine announcement".

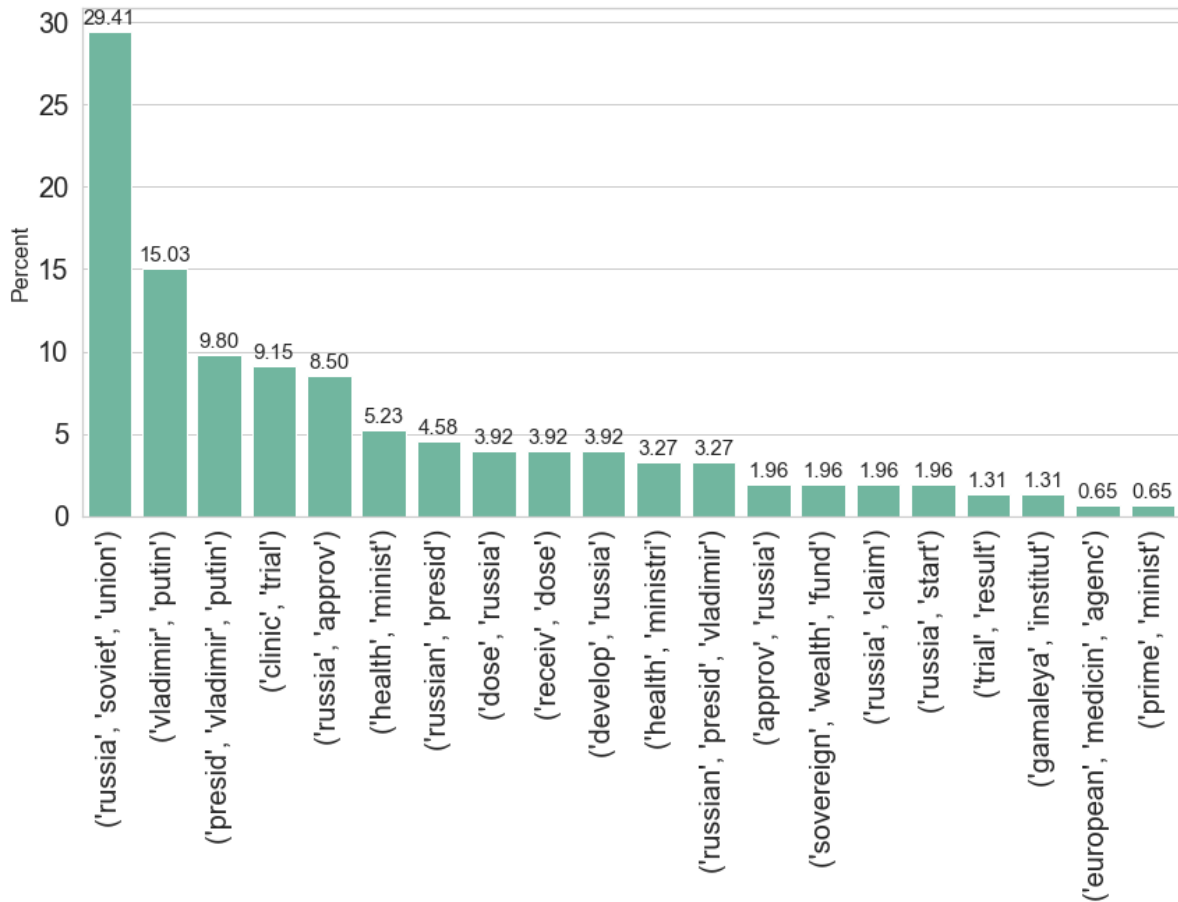
Sputnik V In figures 16 and 17 the 20 most frequent n-grams with respect to Sputnik V for the two periods are illustrated. In contrast to any of the other vaccines, the headlines in both periods have clear indications of the vaccine's origin, in this case Russia. For instance the most frequent n-gram in "Before the COVID-19 vaccine announcement", with a relative frequency of 31.37 percent is "(soviet,union)", while the second highest with almost as many headlines is a trigram containing the most frequent bigram namely "(russia, soviet, union)". Apart from the clear sign of country of origin, Sputnik V in "Before the COVID-19 vaccine announcement" like the other vaccines, have signs of being in a trial period, with the bigrams "(trial, russia)" and "(clinic, trial)" occurring in 11.11 percent and 9.15 percent of the headlines respectively.

Frequent n-grams Sputnik V
 "Before the COVID-19 vaccine announcement"



Supplementary Figure 16: The 20 most frequent bigrams and trigrams in headlines which are about Sputnik for the period "Before the COVID-19 vaccine announcement".

Frequent n-grams Sputnik V "After the COVID-19 vaccine announcement"



Supplementary Figure 17: The 20 most frequent bigrams and trigrams in headlines which are about Sputnik for the period "After the COVID-19 vaccine announcement".

The frequent n-grams for "After the COVID-19 vaccine announcement" illustrate that removing some of the Russian n-grams, to achieve a more nuanced collection of n-grams in both periods, would not have achieved much, as 9 of the 20 most frequent n-grams for "After the COVID-19 vaccine announcement", occur in less than 2 percent of the headlines, with 3 of these being in less than 1 percent. The removal could have been defended with the fact that many of the n-grams for Sputnik V in both periods are somehow repetitions of the same phrases, like with the two most frequent n-grams in "Before the COVID-19 vaccine announcement". A look at the 50 most frequent n-grams with respect to "Before the COVID-19 vaccine announcement", actually showed that the pattern with "Russia" and "trial" related n-grams continue being frequently mentioned after the first 20 illustrated in figure 16.

3. Removal of AstraZeneca from Sentiment analysis

We removed headlines with AstraZeneca coverage from the sentiment analysis for the three periods along with “Vaccine Development” and “Vaccine Rollout” to clarify if AstraZeneca was the main driver in more negatively polarized articles in those periods. The differences between the proportion of negative sentiment headlines including and excluding AstraZeneca in the three periods are illustrated in Supplementary Table 1, while the differences in “Vaccine Development” and “Vaccine Rollout” are illustrated in Supplementary Table 2

	Before COVID-19	Before the COVID-19 vaccine announcement	After the COVID-19 vaccine announcement
With AstraZeneca	57.1 %	36.5 %	39.2
Without AstraZeneca	57.1 %	35.3 %	39.5 %

Supplementary Table 1: *The percentage of vaccine headlines with negative sentiment including and excluding AstraZeneca within “Before COVID-19”, “Before the COVID-19 vaccine announcement” and “After the COVID-19 vaccine announcement”.*

The differences between the three periods are close to identical no matter if AstraZeneca is included or not. The largest difference is in “Before the COVID-19 vaccine announcement” where there are 1.2 % more negative sentiment headlines when AstraZeneca is included.

	Vaccine Development	Vaccine Rollout
With AstraZeneca	22.8 %	60.1 %
Without AstraZeneca	22.4 %	60.1 %

Supplementary Table 2: *The percentage of vaccine headlines with negative sentiment including and excluding AstraZeneca within “Vaccine Development” and “Vaccine Rollout”.*

The small difference in if AstraZeneca is included or not is solely placed in “Vaccine Development”, where there is a difference of 0.4 %.