

## **Supplementary Material for *Quantifying the online news media coverage of the COVID-19 pandemic.***

### **Supplementary Section 1. Data retrieval.**

We retrieved news articles from major international and national online news sources (ONS) in USA, UK, Canada, New Zealand, Australia, Ireland, Russia, Italy, France, Spain and Germany. We selected the ONS in each country by examining the top news providers calculated by website traffic according to SimilarWeb (<http://similarweb.com>) metrics and those that were listed under the BBC media profiles. We were able to confirm internally that numbers from SimilarWeb were accurate by comparing their web traffic estimates to Google Analytics traffic from a major website. This approach to select our media sources increases the likelihood that our ONS would capture a significant portion of the news consumption within each country.

For each ONS we retrieved the landing page snapshots from WebArchive captured since 2015. WebArchive is a reliable online service that specializes in archiving online web content for 477 billion webpages. Since WebArchive frequently captures website snapshots (up to several times per day), we considered it a reliable indicator of the evolution of the news content on each of our ONSs.

For each ONS front-page snapshot extracted from WebArchive we extracted all the links which stayed within the domain of the ONS (see Supplementary Table 1). For each of the links we extracted the raw html, if it was possible directly from the source, otherwise the archived version from WebArchive. If it was impossible to retrieve the link contents either from the original source or the WebArchive, we flagged such links as ‘retrieval error’. The errors in retrieval were defined as pages where the server did not respond correctly, the webpage was unavailable (e.g. 404 errors) or if access to certain resources was blocked (e.g. subscription needed). We curated a global list of error messages after examining the responses for each ONS. In order to assure that we obtain a representative reflection of each ONS, we set the retrieval error rate at 6% (which allowed a good balance between including many ONSs and reliability of link retrieval). There were certain exceptions to this rule where we noted that the links we could not retrieve could not possibly have been article related (e.g. [repubblica.it](http://repubblica.it), [stuff.co.nz](http://stuff.co.nz)). The error rates in retrieval for each ONS are given in Supplementary Table 1.

For each raw html page we extracted the metadata called ‘title’ and ‘description’. Such data elements are typically available for news article content as, among others, they allow for condensed snippets of text to be shared on social networks such as Twitter and Facebook. They are akin in spirit to titles and abstracts of scientific information and serve a similar function of conveying the essence of the document. All things considered, title and description metadata

offer the advantage of being standardized for retrieval, their text length and the condensed information between otherwise heterogeneous for each ONS.

For each ONS, the titles and descriptions were concatenated and associated with all the links pointing to the same text content. This allowed us to collapse url-distinct links pointing to the same content. Such normalization reduces the number of multiple-counts for each article as the raw URL might be updated over several days, or there could be links to comments or video sections all referencing the same core content. The number of such unique links are given in the 'Articles' column in Supplementary Table 1. Furthermore, we analyzed the titles and descriptions associated with the greatest number of links. This allowed us to identify ONS-specific error pages since multiple links that cannot be retrieved linked to the same error title page. Such information was fed backwards into our list of error responses, iteratively refining the retrieval procedure.

Retrieval of all the links from ONS meant that we captured many non-article structural links, such as category pages, terms of use etc. Each title and description text pair was given a per day count as well as count of how many times links pointing to it were observed in all the WebArchive snapshots. Pre-2020 articles provided a large dataset of counts of links that appear very often on ONS front pages, indicating sections etc. Title and description text pairs with per day count above 100 were discarded from the analysis as these were deemed to be structural non-article links.

As the final article filter for each ONS, we performed language classification using the Python library langdetect. Each ONS was associated with a language that articles are expected to be written in (e.g. English for FoxNews and German for Der Spiegel). Articles whose titles and descriptions were not classified as the main language of the ONS were discarded from the analysis. Structural website elements (e.g. section links) or erroneous text was typically language-misclassified (e.g. Daily Telegraph) further reducing the contamination of our dataset with non-article content. In certain cases, such as the BBC, a great number of links were in fact not in English, which is due to the global reach of the service. Certain USA ONS had a sizable proportion of articles in Spanish (e.g. Los Angeles Times or Chicago Tribune). We considered such articles tangential to the overall coverage of these ONS and decided not to include these in the analysis.

The raw front pages, raw individual articles and processed articles are available for download via <http://sciride.org>.

**Supplementary Table 1.** Statistics of retrieval for news media domains. For each domain, we report the number of unique links we were able to identify in the front pages (total links). Out of those links we report how many produced retrieval errors (Number of bad links) and the percentage thereof (%bad links). The Articles column reports the total number of unique articles we extracted from each domain. The Rejected column reports the number of articles that were rejected as being unlikely news articles.

<b>DOMAIN</b>	<b>COUNTRY</b>	<b>TOTAL LINKS</b>	<b>NUMBER OF BAD LINKS</b>	<b>%BAD LINKS</b>	<b>ARTICLES</b>	<b>REJECTED</b>
<b>CTVNEWS.CA</b>	Canada	231,051	99	0	144,582	103
<b>THEGLOBEANDMAIL.COM</b>	Canada	196,001	995	0.5	181,441	222
<b>LAPRESSE.CA</b>	Canada	169,789	2,259	1.3	153,650	237
<b>GLOBALNEWS.CA</b>	Canada	150,027	105	0	130,438	324
<b>ICI.RADIO-CANADA.CA</b>	Canada	147,066	1,613	1	104,264	524
<b>THESTAR.COM</b>	Canada	140,995	417	0.2	134,238	160
<b>MONTREALGAZETTE.COM</b>	Canada	123,027	4,818	3.9	112,947	330
<b>CBC.CA</b>	Canada	103,159	662	0.6	94,849	128
<b>TORONTOSUN.COM</b>	Canada	101,807	1,129	1.1	87,497	340
<b>VANCOUVERSUN.COM</b>	Canada	72,337	2,292	3.1	65,791	243
<b>JOURNALDEMONTREAL.COM</b>	Canada	39,024	155	0.3	35,802	123
<b>MACLEANS.CA</b>	Canada	17,890	72	0.4	17,040	158
<b>RCINET.CA</b>	Canada	11,558	59	0.5	6,661	4,032

<b>TOTAL</b>	Canada	1,503,731	-	-	1,269,200	-
<b>NEWS.COM.AU</b>	Australia	273,231	2,405	0.8	247,873	203
<b>SMH.COM.AU</b>	Australia	250,848	3,926	1.5	227,109	562
<b>THEAGE.COM.AU</b>	Australia	233,318	4,043	1.7	211,739	393
<b>ABC.NET.AU</b>	Australia	177,060	5,205	2.9	149,086	153
<b>AFR.COM</b>	Australia	124,697	517	0.4	116,897	482
<b>9NEWS.COM.AU</b>	Australia	82,263	1,163	1.4	66,585	145
<b>THEWEST.COM.AU</b>	Australia	57,363	208	0.3	56,774	132
<b>SBS.COM.AU</b>	Australia	56,958	822	1.4	48,796	196
<b>TOTAL</b>	Australia	1,255,738	-	-	1,124,859	-
<b>REPUBBLICA.IT</b>	Italy	376,716	44,125	11.7	193,065	771
<b>ILSOLE24ORE.COM</b>	Italy	211,950	5,777	2.7	133,605	1,885
<b>LASTAMPA.IT</b>	Italy	191,831	1,272	0.6	172,802	1,355
<b>HUFFINGTONPOST.IT</b>	Italy	188,637	616	0.3	106,988	625
<b>ILFATTOQUOTIDIANO.IT</b>	Italy	183,423	551	0.3	100,594	304
<b>ANSA.IT</b>	Italy	178,569	288	0.1	169,341	2,994
<b>ILMESSAGGERO.IT</b>	Italy	174,755	1,065	0.6	158,998	174
<b>ILMATTINO.IT</b>	Italy	163,282	386	0.2	129,564	292

<b>CORRIERE.IT</b>	Italy	133,387	974	0.7	105,816	687
<b>TISCALI.IT</b>	Italy	102,778	2,512	2.4	93,442	404
<b>LIBEROQUOTIDIANO.IT</b>	Italy	86,152	5,111	5.9	66,135	126
<b>FANPAGE.IT</b>	Italy	81,871	199	0.2	80,235	302
<b>RAI.IT</b>	Italy	17,892	511	2.8	15,936	377
<b>TOTAL</b>	Italy	2,091,243	-	-	1,526,521	-
<b>DAILYMAIL.CO.UK</b>	UK	1,529,517	49	0	829,047	86
<b>MIRROR.CO.UK</b>	UK	859,725	1,087	0.1	468,632	603
<b>THESUN.CO.UK</b>	UK	791,669	985	0.1	509,583	185
<b>REUTERS.COM</b>	UK	543,163	7,725	1.4	328,651	3,288
<b>THETIMES.CO.UK</b>	UK	486,590	256	0	415,801	65
<b>DAILYSTAR.CO.UK</b>	UK	416,041	226	0	299,874	287
<b>EXPRESS.CO.UK</b>	UK	403,722	385	0	401,076	172
<b>TELEGRAPH.CO.UK</b>	UK	319,967	625	0.1	278,797	2,585
<b>THEGUARDIAN.COM</b>	UK	290,741	45	0	288,567	304
<b>LIVERPOOLECHO.CO.UK</b>	UK	222,963	288	0.1	118,811	261
<b>BBC.COM</b>	UK	215,915	137	0	163,727	26,561
<b>INDEPENDENT.CO.UK</b>	UK	208,125	226	0.1	205,206	281

<b>STANDARD.CO.UK</b>	UK	189,806	4,028	2.1	181,251	479
<b>METRO.CO.UK</b>	UK	170,989	105	0	165,381	86
<b>WALESONLINE.CO.UK</b>	UK	168,752	351	0.2	98,600	373
<b>NEWS.SKY.COM</b>	UK	86,114	1,568	1.8	80,694	79
<b>SCOTSMAN.COM</b>	UK	66,818	1,739	2.6	44,823	125
<b>ITV.COM</b>	UK	63,426	119	0.1	58,714	84
<b>ECONOMIST.COM</b>	UK	24,006	25	0.1	18,314	81
<b>MORNINGSTARONLINE.CO.UK</b>	UK	14,371	435	3	13,636	104
<b>CHANNEL4.COM</b>	UK	8,876	45	0.5	8,607	81
<b>TOTAL</b>	UK	7,081,296	-	-	4,977,792	-
<b>BREITBART.COM</b>	USA	473,460	68	0	175,292	696
<b>LATIMES.COM</b>	USA	428,327	477	0.1	228,673	3,911
<b>WSJ.COM</b>	USA	428,115	333	0	177,647	254
<b>FOXNEWS.COM</b>	USA	418,674	3,988	0.9	289,504	321
<b>WASHINGTONPOST.COM</b>	USA	375,562	2,901	0.7	292,075	431
<b>UPI.COM</b>	USA	370,826	27	0	138,082	72
<b>CHICAGOTRIBUNE.COM</b>	USA	291,095	252	0	146,505	1,948
<b>NEWS.YAHOO.COM</b>	USA	242,964	9,886	4	223,051	224

<b>USATODAY.COM</b>	USA	238,906	1,619	0.6	231,819	113
<b>CBSNEWS.COM</b>	USA	216,257	4,809	2.2	193,922	1,876
<b>BOSTONGLOBE.COM</b>	USA	206,029	3,119	1.5	177,633	135
<b>NYPOST.COM</b>	USA	201,149	36	0	198,003	62
<b>ABCNEWS.GO.COM</b>	USA	200,014	3,050	1.5	172,410	223
<b>NYTIMES.COM</b>	USA	197,711	387	0.1	155,450	187
<b>CNBC.COM</b>	USA	191,100	1,795	0.9	168,541	509
<b>THEHILL.COM</b>	USA	171,544	138	0	170,377	222
<b>BUSINESSINSIDER.COM</b>	USA	145,345	956	0.6	117,468	1,056
<b>NBCNEWS.COM</b>	USA	140,918	266	0.1	126,075	219
<b>WASHINGTONTIMES.COM</b>	USA	139,276	112	0	114,203	292
<b>WND.COM</b>	USA	114,852	381	0.3	78,710	589
<b>TIME.COM</b>	USA	109,595	3,764	3.4	94,575	213
<b>NEWSWEEK.COM</b>	USA	103,173	236	0.2	101,583	493
<b>EDITION.CNN.COM</b>	USA	94,283	2	0	93,871	73
<b>APNEWS.COM</b>	USA	94,259	73	0	92,507	115
<b>SLATE.COM</b>	USA	87,595	562	0.6	64,312	213
<b>POLITICO.COM</b>	USA	83,675	320	0.3	79,311	639

<b>NPR.ORG</b>	USA	70,753	827	1.1	61,537	265
<b>MSNBC.COM</b>	USA	54,395	17	0	53,084	55
<b>THEATLANTIC.COM</b>	USA	49,676	25	0	46,422	182
<b>HUFFPOST.COM</b>	USA	38,524	1,602	4.1	36,119	286
<b>CSMONITOR.COM</b>	USA	37,194	33	0	28,495	148
<b>WESTERNJOURNAL.COM</b>	USA	34,981	686	1.9	31,682	106
<b>INQUIRER.COM</b>	USA	31,029	43	0.1	29,445	229
<b>TOTAL</b>	USA	6,081,256	-	-	4,388,383	-
<b>FRANCETVINFO.FR</b>	France	694,187	726	0.1	573,917	2,675
<b>LEFIGARO.FR</b>	France	507,333	469	0	291,187	718
<b>LEMONDE.FR</b>	France	288,325	3,587	1.2	203,071	1,208
<b>LEXPRESS.FR</b>	France	239,065	143	0	235,163	300
<b>20MINUTES.FR</b>	France	207,226	9,230	4.4	167,383	574
<b>LEPOINT.FR</b>	France	177,423	19	0	171,828	183
<b>LIBERATION.FR</b>	France	166,487	2,002	1.2	149,082	207
<b>LCI.FR</b>	France	113,084	221	0.1	94,273	155
<b>BFMTV.COM</b>	France	66,531	119	0.1	65,704	216
<b>TOTAL</b>	France	2,459,661	-	-	1,951,608	-



<b>WELT.DE</b>	Germany	362,621	980	0.2	307,070	342
<b>SPIEGEL.DE</b>	Germany	313,438	2,247	0.7	270,326	1,749
<b>FOCUS.DE</b>	Germany	276,482	8,907	3.2	250,533	351
<b>FAZ.NET</b>	Germany	266,962	929	0.3	215,398	2,891
<b>N-TV.DE</b>	Germany	251,589	1,506	0.5	239,727	979
<b>HANDELSBLATT.COM</b>	Germany	221,696	5,628	2.5	139,412	391
<b>BILD.DE</b>	Germany	194,428	570	0.2	187,926	610
<b>STERN.DE</b>	Germany	174,504	5,381	3	115,613	4,529
<b>TAGESSPIEGEL.DE</b>	Germany	155,218	130	0	133,925	293
<b>NOZ.DE</b>	Germany	138,794	509	0.3	84,747	447
<b>FR.DE</b>	Germany	117,738	4,856	4.1	64,094	145
<b>ABENDBLATT.DE</b>	Germany	109,148	5,794	5.3	98,415	170
<b>TAGESSCHAU.DE</b>	Germany	104,377	3,021	2.8	50,985	88
<b>MORGENPOST.DE</b>	Germany	100,603	6,317	6.2	87,741	188
<b>BZ-BERLIN.DE</b>	Germany	55,698	202	0.3	54,690	250
<b>NDR.DE</b>	Germany	50,989	10	0	10,640	93
<b>WDR.DE</b>	Germany	40,966	1,306	3.1	18,667	332
<b>RTL.DE</b>	Germany	20,105	29	0.1	18,494	77

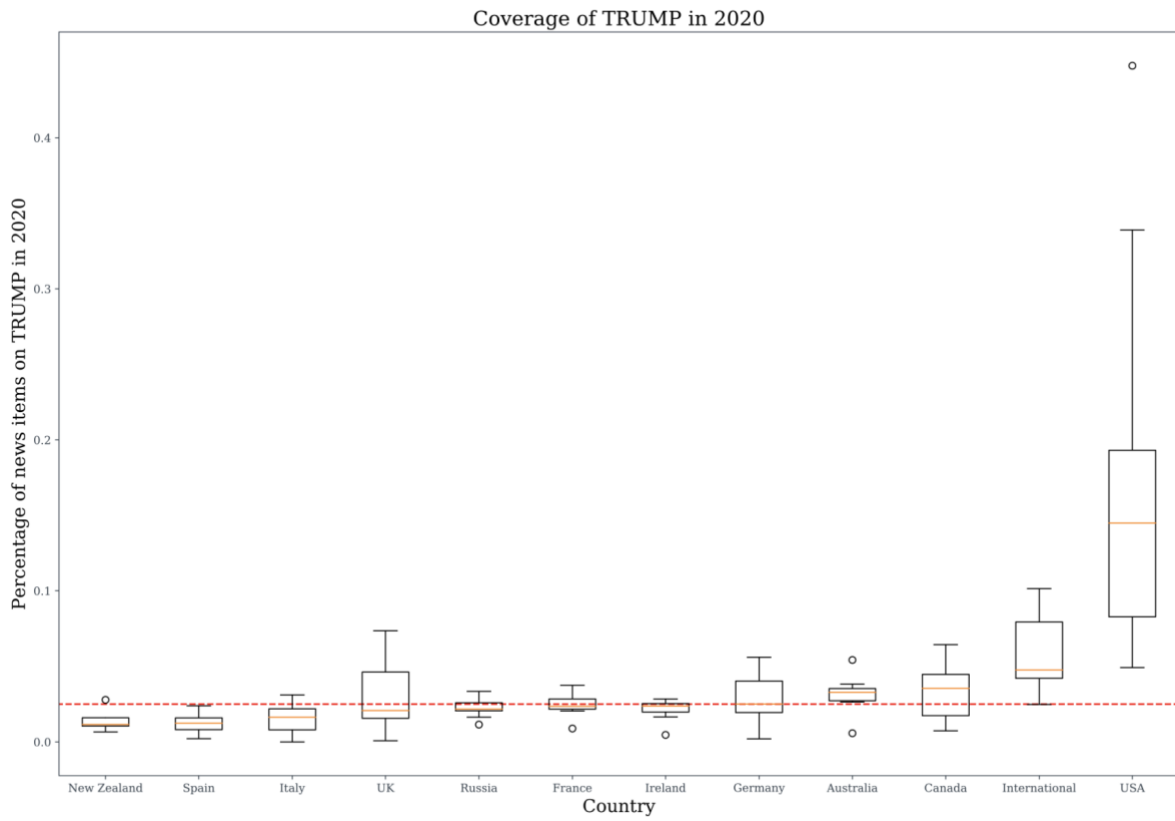
<b>TOTAL</b>	Germa ny	2,955,356	-	-	2,348,403	-
<b>THESUN.IE</b>	Ireland	196,122	522	0.2	161,165	29
<b>INDEPENDENT.IE</b>	Ireland	191,800	5,250	2.7	179,924	577
<b>THEJOURNAL.IE</b>	Ireland	178,462	222	0.1	102,461	72
<b>IRISHTIMES.COM</b>	Ireland	176,304	103	0	161,235	393
<b>IRISHEXAMINER.COM</b>	Ireland	140,383	7,523	5.3	126,892	282
<b>IRISHMIRROR.IE</b>	Ireland	104,582	311	0.2	66,671	64
<b>BREAKINGNEWS.IE</b>	Ireland	78,338	234	0.2	76,423	82
<b>RTE.IE</b>	Ireland	32,530	51	0.1	30,827	127
<b>TOTAL</b>	Ireland	1,098,521	-	-	905,598	-
<b>EURONEWS.COM</b>	Interna tional	138,093	4,553	3.2	129,879	168
<b>DW.COM</b>	Interna tional	112,242	2,599	2.3	100,029	278
<b>RT.COM</b>	Interna tional	103,067	1,574	1.5	99,282	63
<b>ALJAZEERA.COM</b>	Interna tional	66,785	34	0	65,477	458
<b>FRANCE24.COM</b>	Interna tional	58,714	710	1.2	57,005	49
<b>RFI.FR</b>	Interna tional	11,550	42	0.3	11,317	34
<b>TOTAL</b>	Interna tional	490,451	-	-	462,989	-

<b>STUFF.CO.NZ</b>	New Zealand	276,231	22,793	8.2	221,823	641
<b>NZHERALD.CO.NZ</b>	New Zealand	274,618	3,703	1.3	245,787	515
<b>SCOOP.CO.NZ</b>	New Zealand	124,309	88	0	119,624	708
<b>NEWSHUB.CO.NZ</b>	New Zealand	49,696	2,141	4.3	46,817	95
<b>RNZ.CO.NZ</b>	New Zealand	18,944	1,041	5.4	16,999	77
<b>TOTAL</b>	New Zealand	743,798	-	-	651,050	-
<b>RBC.RU</b>	Russia	382,762	751	0.1	246,376	255
<b>RIA.RU</b>	Russia	352,608	62	0	282,422	206
<b>KOMMERSANT.RU</b>	Russia	318,046	314	0	232,298	289
<b>GAZETA.RU</b>	Russia	283,339	80	0	240,404	566
<b>LENTA.RU</b>	Russia	267,014	508	0.1	263,490	284
<b>INTERFAX.RU</b>	Russia	252,237	194	0	228,483	404
<b>RAMBLER.RU</b>	Russia	245,294	1,622	0.6	206,648	338
<b>MK.RU</b>	Russia	245,201	230	0	239,912	453
<b>NTV.RU</b>	Russia	224,220	377	0.1	136,400	797
<b>AIF.RU</b>	Russia	216,288	21	0	172,920	345

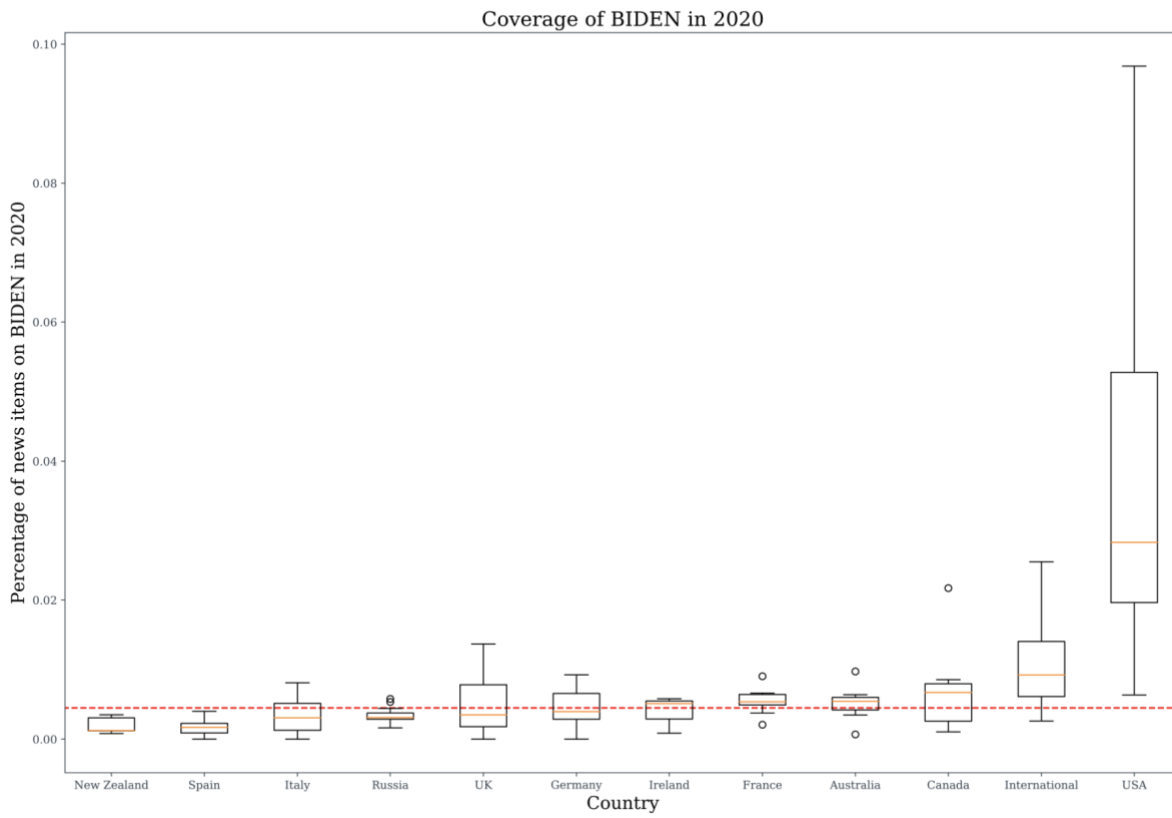
<b>VESTI.RU</b>	Russia	188,944	1,527	0.8	168,850	200
<b>IZ.RU</b>	Russia	179,549	15	0	165,034	346
<b>KP.RU</b>	Russia	157,505	297	0.1	152,472	381
<b>TASS.RU</b>	Russia	153,497	945	0.6	151,541	574
<b>RG.RU</b>	Russia	142,451	75	0	139,365	279
<b>NEWSRU.COM</b>	Russia	105,902	496	0.4	100,704	53
<b>TVZVEZDA.RU</b>	Russia	90,639	4,364	4.8	84,260	348
<b>NG.RU</b>	Russia	86,974	3,341	3.8	66,217	329
<b>REN.TV</b>	Russia	72,092	100	0.1	71,029	662
<b>TOTAL</b>	Russia	3,964,562	-	-	3,348,825	-
<b>ABC.ES</b>	Spain	653,473	2,448	0.3	330,934	3,378
<b>COPE.ES</b>	Spain	419,512	3,678	0.8	242,826	3,332
<b>20MINUTOS.ES</b>	Spain	404,122	1,535	0.3	248,585	762
<b>ELMUNDO.ES</b>	Spain	355,466	636	0.1	201,989	1,198
<b>ELPAIS.COM</b>	Spain	311,945	350	0.1	200,784	636
<b>CADENASER.COM</b>	Spain	297,416	1,224	0.4	251,223	9,043
<b>ELPERIODICO.COM</b>	Spain	292,584	1,251	0.4	221,135	240
<b>EUROPAPRESS.ES</b>	Spain	287,886	5,235	1.8	276,158	1,296

<b>LARAZON.ES</b>	Spain	268,640	7,565	2.8	166,949	304
<b>ELESPANOL.COM</b>	Spain	231,377	4,429	1.9	137,183	1,668
<b>LAVANGUARDIA.COM</b>	Spain	199,272	681	0.3	194,336	1,508
<b>HUFFINGTONPOST.ES</b>	Spain	190,885	629	0.3	109,330	1,844
<b>ELDIARIO.ES</b>	Spain	165,130	405	0.2	124,094	3,222
<b>RTVE.ES</b>	Spain	132,005	361	0.2	108,166	103
<b>PUBLICO.ES</b>	Spain	131,274	335	0.2	111,370	4,921
<b>TELECINCO.ES</b>	Spain	64,818	2,849	4.3	58,657	201
<b>ANTENA3.COM</b>	Spain	52,484	132	0.2	49,916	236
<b>ONDACERO.ES</b>	Spain	50,935	88	0.1	49,604	196
<b>CUATRO.COM</b>	Spain	42,919	738	1.7	39,472	413
<b>TOTAL</b>	Spain	4,552,143	-	-	3,122,711	-
<b>TOTAL</b>	Overall	34,277,756	-	-	26,077,939	-

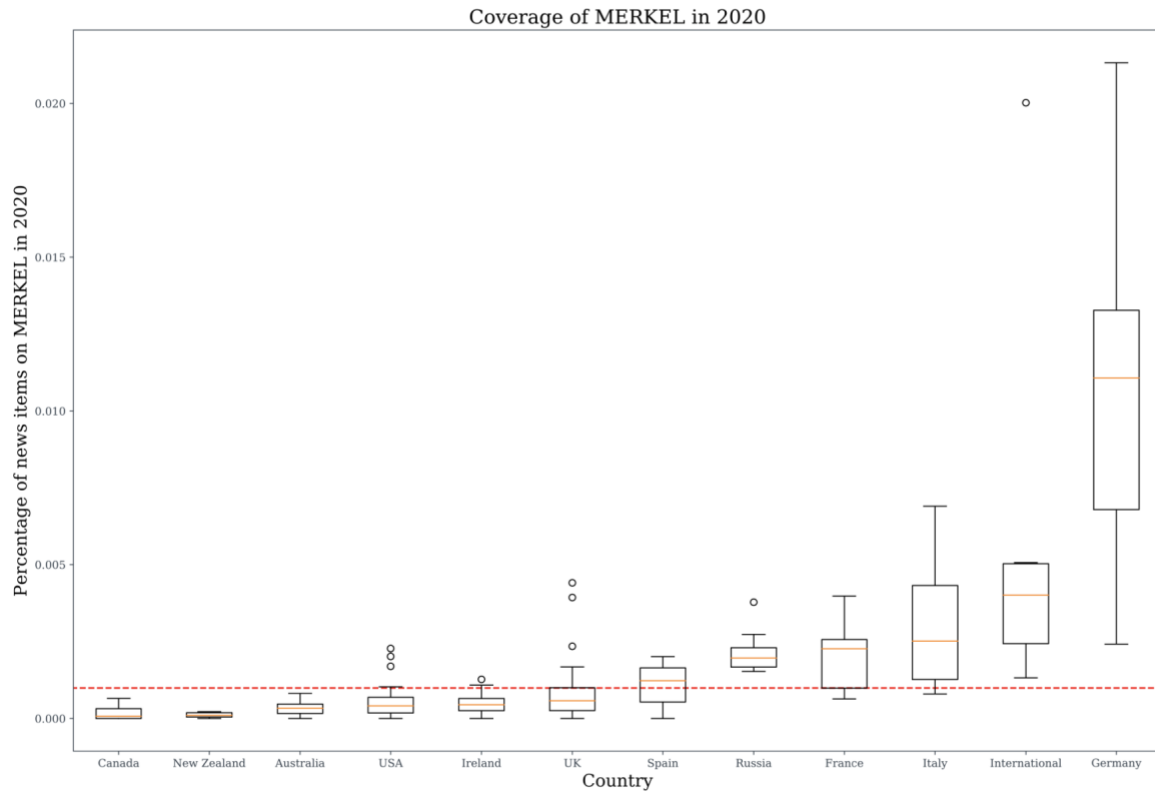
## Supplementary Section 2. Coverage.



Supplementary Figure 1. Volume of coverage of Donald Trump in 2020.

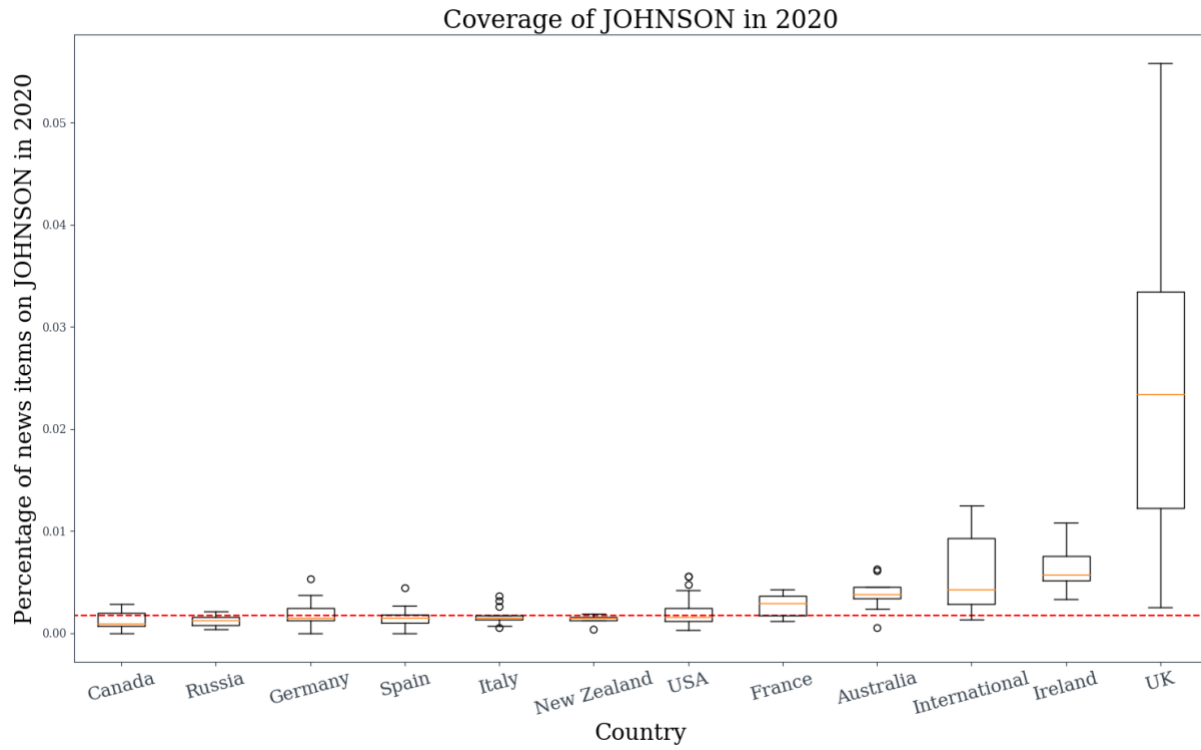


**Supplementary Figure 2.** Volume of coverage of Joe Biden in 2020.

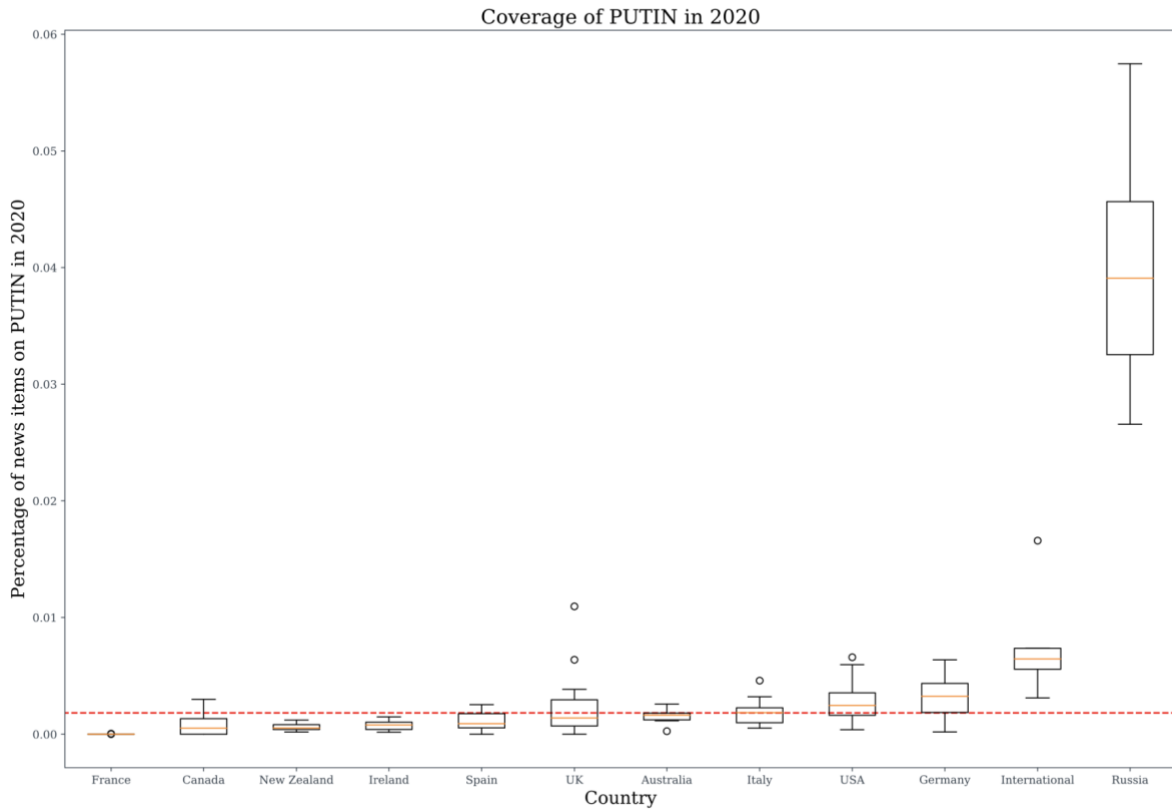


**Supplementary Figure 3.** Volume of coverage of Angela Merkel in 2020.





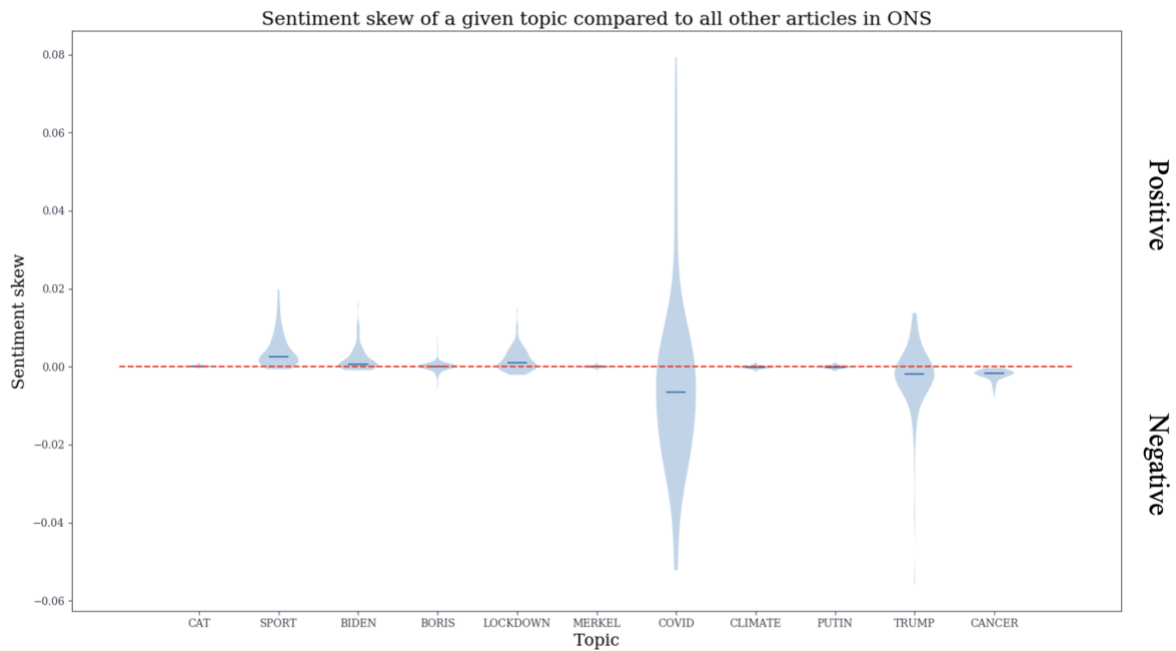
**Supplementary Figure 4.** Volume of coverage of Boris Johnson in 2020.



**Supplementary Figure 5.** Volume of coverage of Vladimir Putin in 2020.

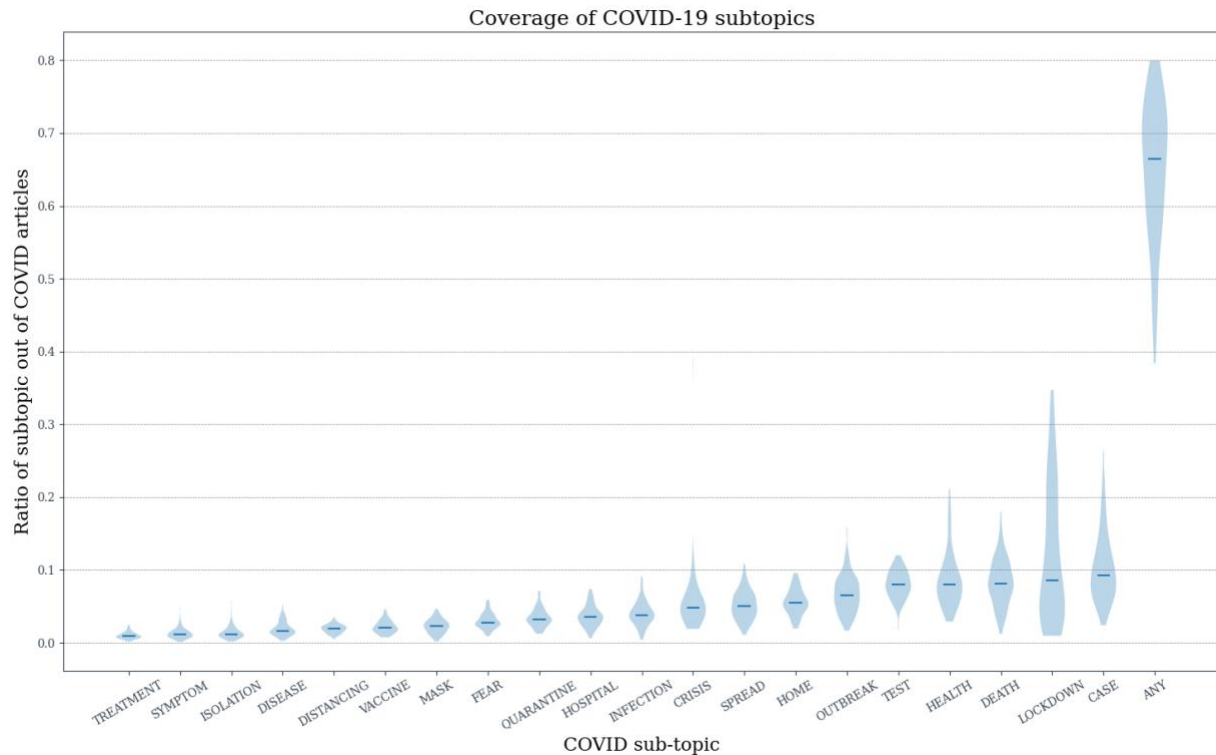
### **Supplementary Section 3. Sentiment.**

Topics such as *CANCER* however, an order of magnitude more negative, receive much less coverage than *COVID-19*. Out of 2,149,585 front page articles from English-speaking ONSs in 2020 we identified 589,701 as *COVID* and 9,548 as *CANCER* (Table 3). The volume of the coverage for a given topic might play a role in being able to polarize the overall sentiment perception. To illustrate this, in supplementary Figure 6 we plotted the sentiment skew as in Figure 2, but with individual ONS sentiment differences weighted by the proportion of topic coverage. The sentiment skew becomes more pronounced for topics which received a large amount of media coverage (e.g. *COVID*) than for topics with less media coverage (e.g. *CANCER*). This intuitively illustrates the higher contribution to sentiment polarization for topics with relatively lower negativity but high coverage volume (e.g. *COVID*) than more negative topics that receive less coverage (e.g. *CANCER*).

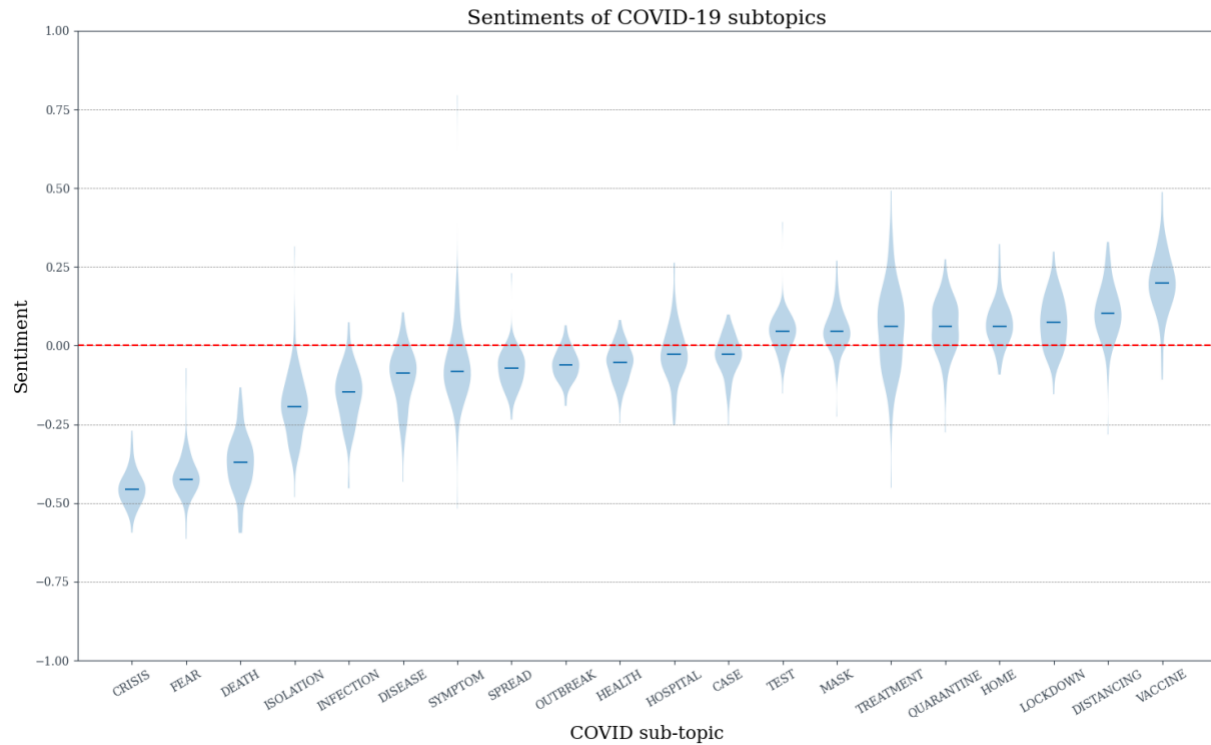


**Supplementary Figure 6.** Relative Sentiment Skew of news in 2020, weighted by percentage volume coverage in 2020. The rskew formula was adjusted by the proportion of coverage of a given topic as  $(\mu_{ONS, TOPIC} - \mu_{ONS, TOPIC}) * coverage_{ONS, TOPIC}$  where  $coverage_{ONS, TOPIC}$  is the ratio of articles on a specific topic in 2020 in a specific ONS.

## Supplementary Section 4. COVID-19 Sub-topics.



**Supplementary Figure 7. Subtopic coverage per ONS.** For each of 91 English-speaking ONS we calculated the proportion of all the COVID articles that a given subtopic accounted for. The violin plots are aggregates of such proportions per-subtopic from each of 91 ONS. The subtopic ‘ANY’ denotes the proportion of articles that can be classified as one of the other subtopics in the Figure.



**Supplementary Figure 8. Subtopic sentiment per ONS.** For each of 91 English-speaking ONS we calculated the mean VADER sentiment of all the articles that a given subtopic accounted for. The violin plots are aggregates of such mean sentiments per-subtopic from each of 91 ONS.

**Supplementary Section 5. Estimating sentiment polarization after removing a number of articles.**

We tested whether removing all articles identified as specific topics (e.g. COVID-19) from ONS would result in a statistically significant polarization towards either positive or negative sentiment. We noted the mean sentiment after removing all articles on a given topic from a specific ONS. We then assessed how extreme the given sentiment mean was by approximating its distribution by repeatedly removing the same number of random articles from the outlet.

Let  $ONS_{all}$  denote all articles in a given ONS,  $ONS_{TOPIC}$  all articles not identified as topic in a given ONS and  $ONS_{TOPIC}$  all articles on a given topic in the given ONS. We noted the mean sentiment of 2020 articles when a given topic was removed  $\mu_{ONS, TOPIC}$ . We then sampled  $|ONS_{TOPIC}|$  articles from  $ONS_{all}$  and noted the sentiment of the remaining articles denoted  $\mu_{ONS, sample}$ . Each such sample provided an indicator of possible average sentiment when the same number of articles as  $|ONS_{TOPIC}|$  is sampled. For each ONS the procedure of producing  $\mu_{ONS, sample}$  was repeated  $\min(1000, |ONS_{TOPIC}|)$  times, approximating the distribution of sentiment means upon removal of the given number of articles. We noted the number of times

$\mu_{ONS, \text{TOPIC}}$  was higher than  $\mu_{ONS, \text{sample}}$  as  $H_{ONS}$  and the number where it was lower as  $L_{ONS}$ . Large  $H_{ONS}$  value indicates that removing TOPIC articles from ONS on average results in sentiment polarization towards positive, as compared to random sampling the same number of articles from ONS and vice versa for  $L_{ONS}$ . In both cases we set the significance level at 5% ( $H_{ONS} > 0.95\%$  and  $L_{ONS} < 0.05$ ). Since we had 91 English-speaking ONSs, we corrected for multiple testing dividing by 91 (the Bonferroni correction). For each ONS and a given topic we therefore reported whether removing a topic had a significant polarization towards either positive or negative sentiment ( $H_{ONS} > 0.95/91$  and  $L_{ONS} < 0.05/91$ ) or whether no determination could be made ( $H_{ONS} < 0.95/91$  and  $L_{ONS} > 0.05/91$ ).