

WP5, juni 2020: Social Media Monitoring

Fase 4: Twitter-analyse *tests, mondkapjes, vragen*

5.1.2e, RIVM; 5.1.2e, 5.1.2e,
 LIACS (Universiteit Leiden); 5.1.2e
 5.1.2e, CLST (Radboud Universiteit); 5.1.2e, KNAW Meertens Instituut; 5.1.2e
 5.1.2e Netherlands eScience Center

Fact sheet: belangrijkste observaties

Van de vier centrale vragen van de gedragsunit richten we ons op vraag 3 (*Draagvlak voor de maatregelen*) en 4 (*Vertrouwen over de besluitvorming ten aanzien van maatregelen en plan van aanpak door Rijksoverheid*). We analyseerden tweets over **coronatests** en **mondkapjes**.

- Tests:
 - De Twitterberichten over testen bevatten de meeste negatieve emoties tussen 22 en 29 mei. Na 1 juni krijgen de positieve berichten de overhand. De meest voorkomende positieve woorden over testen suggereren **draagvlak voor testen** ('controle', 'betrouwbaar'). Er wordt in een klein aantal berichten geklaagd dat mensen pas laat terecht kunnen voor een test (~8%) en dat de tests niet nauwkeurig zijn (~2%).
 - De aandacht voor verschillende *aspecten* van tests verschuift. Tussen februari en juni 2020 daalde de aandacht voor **testresultaten** (30% in februari, 18% in mei) en steeg die voor **testmaterialen**. **Testbeleid** beleefde een piek aan aandacht in maart.
- Mondkapjes:
 - In Twitterberichten over mondkapjes zien we twee periodes met veel negatieve emoties: rond 15 mei en rond 1 juni. In beide periodes wordt **wantrouwen in het (onduidelijke) mondkapjesbeleid** van de overheid gedeeld.
 - Vooral oudere, mannelijke gebruikers tweeten over mondkapjes, zij focussen vooral op de onderwerpen **beschikbaarheid** en **verplichting** tot dragen van mondkapjes.

Op Twitter worden ook veel vragen gesteld. Met clusteranalyse groepeerden we gestelde vragen in coherente clusters over hetzelfde onderwerp. Er worden doorlopend veel vragen gesteld over de **gevolgen voor de economie** en er is voortdurend **onduidelijkheid over de interpretatie van maatregelen**, en **kritiek op de onlogica** ervan. Het onderwerp **complottheorieën** verdween in mei uit de top-4, terwijl steeds meer mensen vragen stellen over **aanpassingen** naar het nieuwe normaal. Andere terugkerende vragen richten zich op de interpretatie van **statistieken**, **medische aspecten** van de ziekte, het plannen van **vakanties**, en de angst voor een **tweede golf**.



Analysis of data about corona tests

Datasets:

- Twitter: 3,466 tweets, 1 - 22 June 2020
- Nu.nl: 6,771 user comments, 1-15 June

Most discussed topics in June (testing)

These topics are the most shared bits of information/opinions.

Twitter

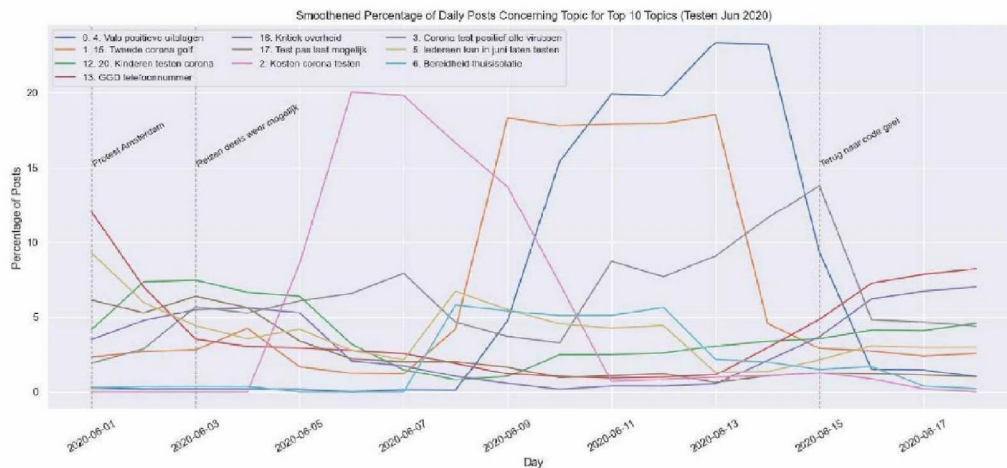


Nu.nl



Some of the Twitter topics are criticizing the Dutch government:

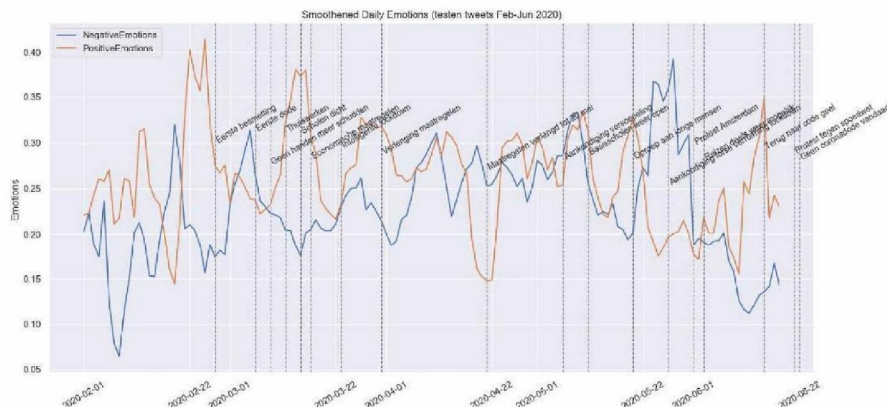
To pic #	Label	% of tweets	Top 10 woorden
4	Vals positieve uitslagen	~8%	marges vage negatives reacties false rivm momentopname onbetrouwbaar blamage meestgebruikte
6	Bereidheid thuisisolatie	~2%	bereidheid thuisisolatie grote verkoudheidsklachten blijven mensen positieve test operatie ondergaan
16	Kritiek overheid	~2%	goed blundert verzorgingshuizen heen mondkapjes crisis vind corona test gedachten
17	Test pas laat mogelijk	~2%	horen thuis gerelateerd collega krijgt klachten corona test afsprakennummer 0800



Prevalence of topics related to testing on Twitter over time (June). Three topics have generated many tweets during the month of June. The costs of testing were a much discussed topic in the beginning of the month. Subsequently, the second wave was the most important topic for a number of days. The most recent topic that was discussed in a large number of tweets were the false positive tests. Currently Tweets about the telephone number of the GGD are increasing.

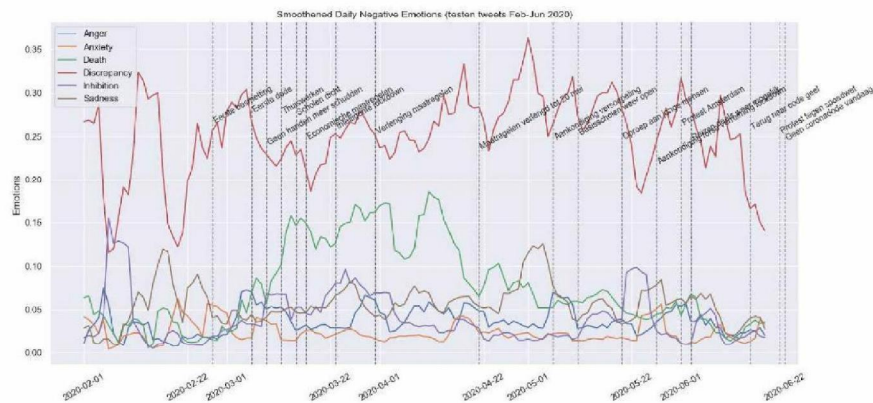
Emotions over time since February (tests, Twitter)

(proportions of words related to a LIWC category)



Patterns we see:

- Negative emotions peak between 22 and 29 May. This is the period in which the lockdown came to an end (and testing became more important). Following that peak, the positive emotions took over after June 1st.
- The most frequent positive words related to testing are: positief, goed, beter, controle, dank, hoop, vrij, fijn, lekker, betrouwbaar. This suggests that the positive emotions here are at least partly related to trust in the testing policy
- The most frequent negative words related to testing are: alleen doden protest gebrek protesteren geleden onvoldoende erg strijd isoleren. Here we clearly see a reference to the protests that took place in the beginning of June. Isolating is actually not a negative word in this context because it relates to the policy of 'test/trace/isolate'.



Zooming in on the negative emotions, we see:

- Death follows the same trend as the number of reported deaths
- Discrepancy scores high throughout the whole period, peaking in the last week of May. The top words for this category are: maar, wil, moet, moeten, zou, willen, zal, zouden, zullen, hoop
- The other negative emotions are quite stable in the time period

Word categories related to corona tests

LIWC categories that are at least 1.3 times more frequent in tweets about coronatesting than in random tweets from 2019

Category	Top-10 words
Inhibition	stoppen voorkomen beperkt verhinderen controle verboden beperken verbiedt beperking plicht
Sport	beweging golf skiën spelen speelt oefening conditie voetbal spel sport
PersonalCare	wassen zeep bad schoon haren poetsen waste shampoo schoonmaken wast
Humans	mensen personen kinderen kind volk team partner mens vrouwen mannen

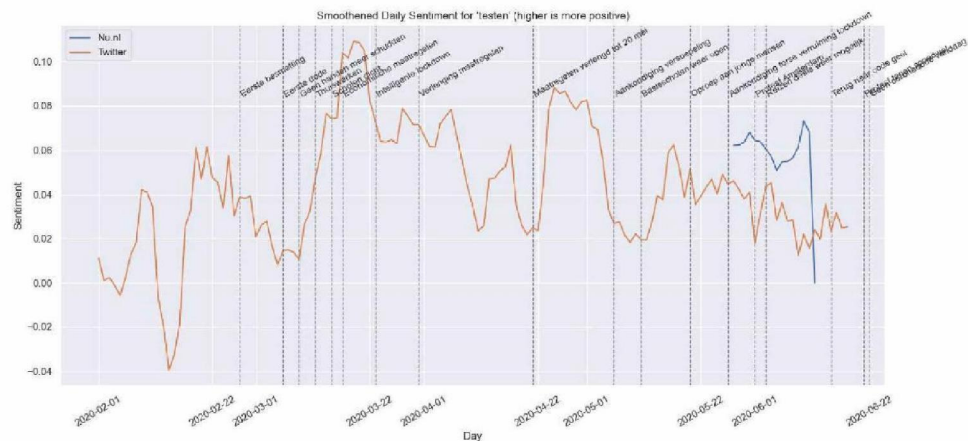
Uit deze tabel blijkt dat Inhibition of remming een veelvoorkomend psychologisch construct is op twitter. De Twitteraars plaatsen veel berichten over alles wat niet mag. De categorieën sport en persoonlijke verzorging laten zien dat mensen zich bezig houden met hoe ze zichzelf door de beperkingen heen kunnen slaan en in goede gezondheid kunnen blijven. Ook is persoonlijke verzorging gerelateerd aan de coronamaatregel om handen te wassen.

LIWC categories that are at least 1.3 times less frequent in tweets about coronatesting than in random tweets from 2019

Category	Top-10 words
See	lees zien zie kijk gezien kijken ziet bekijk kijkt visie
Assent	ja zeker juist orde inderdaad prima goedgekeurd overeenkomst oké accepteren
TelevisionMovies	video film reclame televisie soap bioscoop zender acteur toneel vertoning

Sentiment over time since February (testing)

Estimated sentiment in tweets (higher is more positive)



Two interesting peaks in positive utterances:

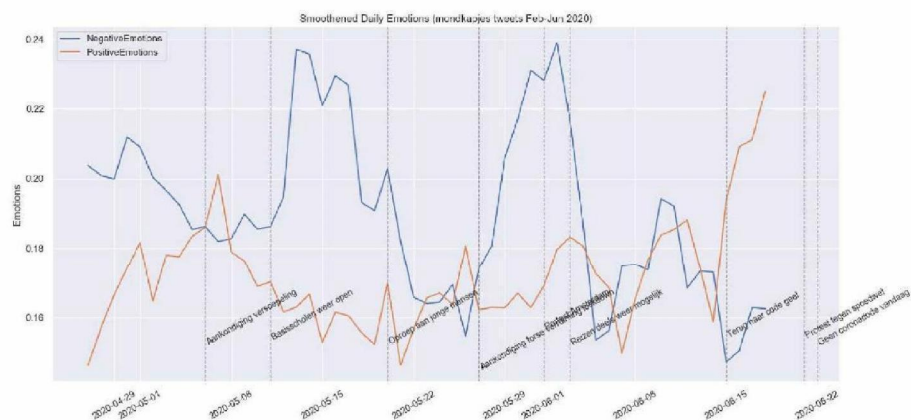
- Between 10 and 20 March. This was the period when the schools closed and everyone started working from home.
- Between 22 and 27 April. This is the period in which the regulations were extended until May 20.

Analysis of data about mondkapjes

Dataset:

- Twitter: 192,942 tweets, 27 April - 14 June 2020

Emotions over time (mondkapjes)



The negative emotions show two clear peaks in the mondkapjes sample:

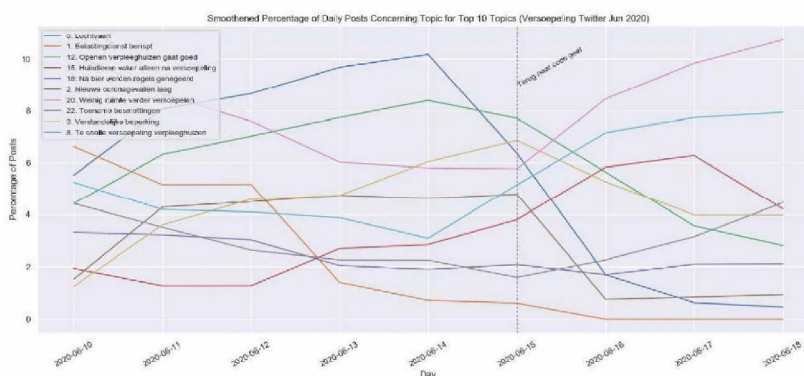
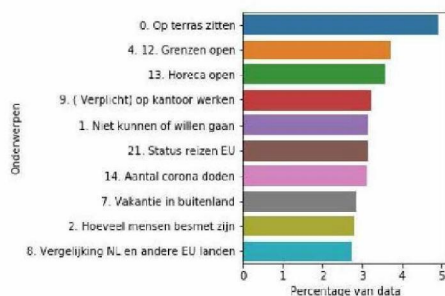
- Around May 15: utterances about distrust in the government (“schippers blijft bullshit verkopen”, “het maandenlange falen van #jaapvandissel en #rivm”; “nederland discussieert al 3 maanden over #mondmaskers of #mondkapjes”), the uselessness of mondkapjes (schijnveiligheid, nutteloos, fopkapjes), and discomfort (onhandig)
- Around June 1st: utterances about distrust in the government (“de klucht met de #mondkapjes blijft doorgaan. verkoop medische mondkapjes in nederland verboden”), and practical complaints about the use of mondkapjes (“het is 1 juni. maar toch laten de bestuurders klanten zonder mondkapjes binnen in de bus.”)

The positive emotions started to peak in the last week. Many tweets in these last week are about the protests, and the positive emotions related to mondkapjes are relief and praise that people wear mouth masks during the protests.

Analysis of data about versoepeling

Dataset:

- Nu.nl: 5,420 user comments, 1-14 June
- Twitter: 1,750 tweets, 8-22 June



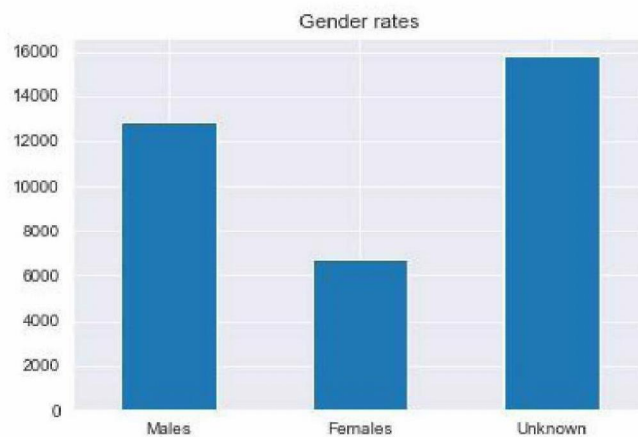
Roze lijn (topic 20): RIVM geeft aan dat er weinig ruimte is om verder te versoepelen. Dit wordt een belangrijk onderwerp na de versoepeling op 15 juni. Veranderingen in de luchtvaart (topic 0, donkerblauw) wordt sinds midden Juni minder over gesproken. Tegelijkertijd lijkt er wel steeds meer kritiek op snelle opening verpleeghuizen (topic 8, lichtblauw)

Twitter Demographics

We intend to investigate who are tweeting about face masks. This analysis focuses on automatic estimations of the demographic characteristics of the twitter population in the samples that form the basis of our reports. This population analysis provides a background or a framework when interpreting the content of the tweets. We performed our analysis of the author profiles of the tweets (twitter users) in the sample that was collected between 27 April 2020 and May 25, 2020.

Gender

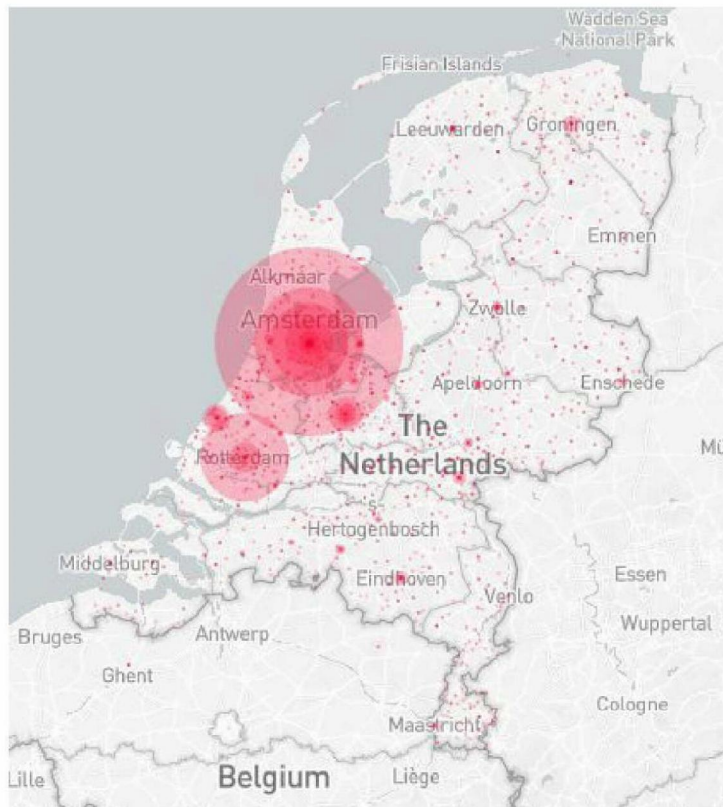
We determined the gender of Twitter users by looking up user names in a male and female Dutch first name list from the Meertens Institute. Names without a match are listed as unknown, and are likely not first names but often organization names. We found twice as many males as females in our sample.



Gender prediction of Twitter users based on first names in profile names

Locations

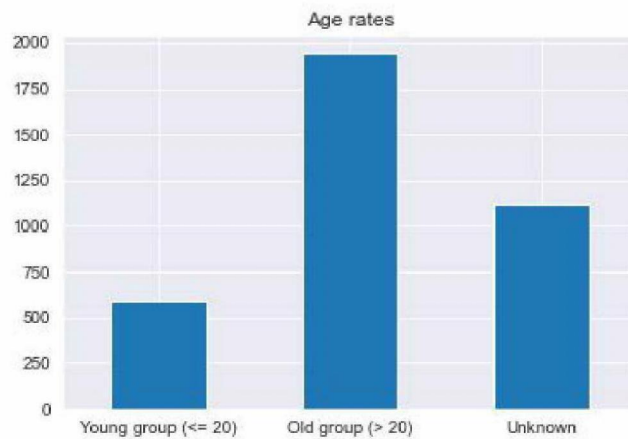
Around two-thirds of all Twitter users indicate a location description in their author profile. We normalized these descriptions and mapped them to a geographical map as shown below. Tweets about face masks originate from all regions in the Netherlands. Locations in Belgium are listed in less than 1% of the accounts. A majority lists their location simply as 'the Netherlands', followed by indications of major cities (Amsterdam, Rotterdam, Utrecht). Note that Amsterdam is listed as location three times more often than Rotterdam and about ten times more often than Eindhoven. The dots in the image below are scaled values, not actual counts.



Overview of locations mentioned in twitter user profiles. Increased dot size and color intensity indicate a higher number of users in that location.

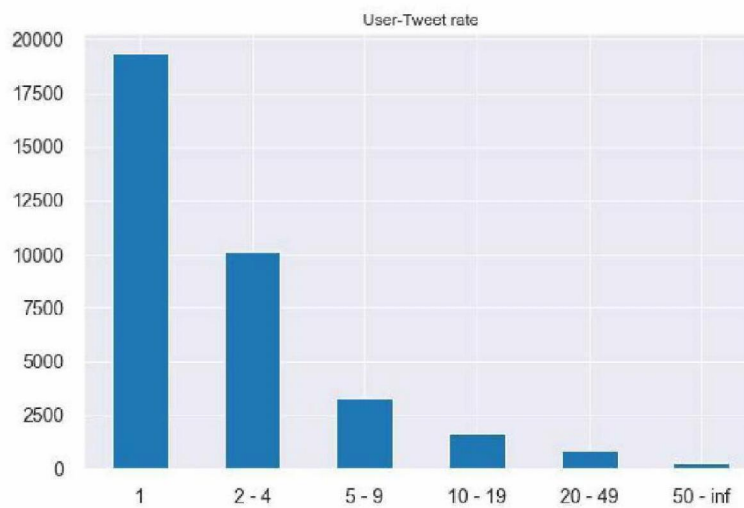
Age

Most twitter user profiles do not list the age of the users and estimating the age of the user is not a trivial task (Nguyen et al, 2013). We can get a crude indication of age indications by applying supervised classification to a tweet sample from each user profile. As we have thousands of Twitter users to investigate, we have chosen a simplified approach. Each twitter user is represented by his/her last 100 tweets and we predict whether the user is younger than 20 or 20 and older. We use a bag-of-words approach and a Naive Bayes classifier. Only when the class label probabilities are at least .2 different from each other to exclude uncertain cases where both ages seem to be likely according to the classifier. Due to time limitations, we report on a subset of the twitter users. Our findings show that the older twitter users are the vast majority.



Twitter behavior

We present some statistics on the number of tweets, friends and followers in the sample. Most users only tweet once about face masks, and very few (42) users tweet over fifty times. We observe that on average users have more followers than friends (people they follow).



Number of tweets per twitter user in the sample

	Favourites	Followers	Friends	Tweets
count	35393	35393	35393	35393
mean	8985.78	1916.32	755.94	3.7
std	23282.81	22794.54	1710.06	8.92
min	0	0	0	1
25%	359	71	136	1
50%	2000	248	347	1
75%	7846	794	823	3
max	796977	2339626	110478	407

Average number of endorsements (favourites), followers and friends (following) in the face mask sample.

Gender and age: analysis of the face masks sample

We have taken the tweets in the face mask sample for which we had age and gender predictions and split the sample in old versus young, and male versus female.

These samples were analyzed with Pinto (Oostdijk et al. 2019) with the same lexicon and linguistic patterns as was described in the WP5 report of 27 May, 2020. The lexicon and patterns summarize the topics that are being discussed when tweeting about face masks.

When comparing the topics discussed by older versus younger twitter users, we notice that older users are more concerned with the availability of face masks and the obligation to wear them. Young users tweet more about the spreading of the virus and voice their opinions.

	all-O		all-Y	
	#	%	#	%
beschikbaarheid	512	13.8	51	9.4
verplichting	1469	39.7	194	35.9
protectie	820	22.2	123	22.8
verspreiding	400	10.8	80	14.8
nut-noodzaak	92	2.5	21	3.9
elders	29	0.8	6	1.1
opinie	142	3.8	39	7.2
geschiktheid	125	3.4	9	1.7
risico	107	2.9	17	3.1
praktisch	3	0.1	0	0.0
unspecified	1	0.0	0	0.0
	3700	100	540	100

Figure shows the Pinto analysis on old (O) versus young (Y) users and the differences of the distribution of the discussed topics.

This analysis shows that males users are more concerned with the availability of face masks and the obligation to wear them than females, while females tweet more about protection by face masks and prevention of spreading of the virus.

	all-F	all-F	all-M	all-M
	#	%	#	%
beschikbaarheid	297	14.2	644	16.3
verplichting	735	35.1	1443	36.4
protectie	551	26.3	834	21.1
verspreiding	239	11.4	445	11.2
nut-noodzaak	62	3.0	114	2.9
elders	18	0.9	48	1.2
opinie	64	3.1	196	4.9
geschiktheid	52	2.5	119	3.0
risico	78	3.7	117	3.0
praktisch	0	0.0	0	0.0
unspecified	1	0.0	0	0.0
	2097	100	3960	100

Figure shows the Pinto analysis on female (F) versus male (M) users and the differences in the distribution of the discussed topics.

In-depth analysis of 'test' tweets

In this part of our study we investigated what users on Twitter were tweeting about as regards corona testing between 1 February and 1 June 2020. The data used had been collected using 'test' as keyword. For this study all data were included (i.e. no duplicates or near duplicates were removed). The data were organized in 4 data files (collections), one for each month based on the date of publication. The sizes of the collections varied from 7,110 tweets for the month of February up to as many as 132,178 tweets for March 2020 (Table 1):

Table 1. Collection sizes

Collection name	month	# of tweets
202002	Feb 2020	7,110
202003	Mar 2020	132,178
202004	Apr 2020	86,340
202005	May 2020	48,063

The analysis was performed by means of Pinto. Whenever the system matched one of the specified linguistic patterns it would return the associated label. The set of labels was as follows:

1. Testen
2. Testresultaten
3. Testmaterialen
4. Testbeleid
5. Testcapaciteit
6. Doelgroep
7. Testmethode
8. Testfaciliteiten
9. Testgelegenheid – elders – kosten (originally three separate labels, combined for reporting purposes)

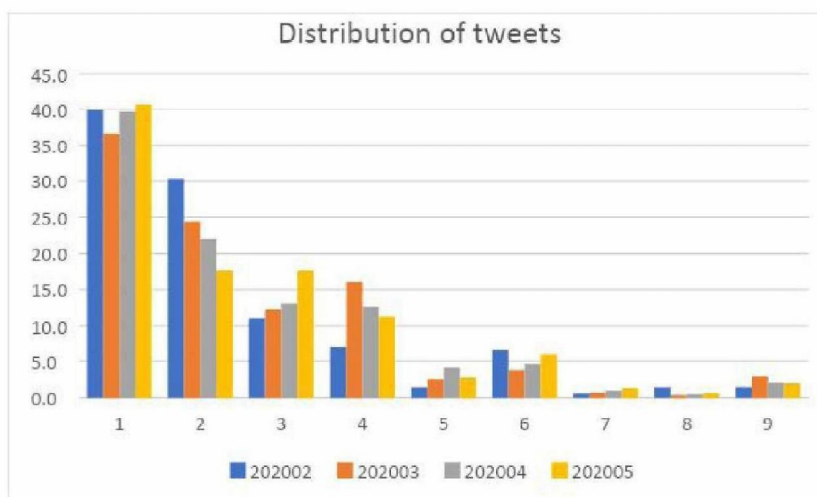


Figure 1. Distribution of 'test' tweets over different topic classes

The results (Figure 1) are very much in line of what one might have expected:

- In February the corona crisis was still very much something that was happening mostly elsewhere, until there was a German person just across the border that tested positive for corona and the first Dutch cases were reported. During this month 'testresultaten' were a frequent topic, the attention for which decreased over time (from 30.3% in Feb to 17.6% in May 2020).

This is also reflected in the top-5 tweets for this month, where 4 out of 5 tweets are about testresults:

freq	Tweet text
423	"Vanavond werd bekend dat een Duitse man positief testte op COVID-19. De man verbleef een week eerder in Limburg. De GGD is gestart met contactonderzoek. Mochten deze contacten klachten krijgen, dan worden zij geïsoleerd en getest. Morgen volgt een update. https://t.co/D3FZTtv03X https://t.co/eYExaKz9n4 "
127	"#Baudet had 10 dagen geleden een spoeddebat aangevraagd iz #coronavirussus. Dit werd rigoreus naar de prullenbak verwezen. Blok hoefde niet getest te worden, heeft iemand hem nog gesignaleerd? Nu is het stil, eng stil. Iemand in Den Haag?"
116	Vannacht testten de partner en het jongste van de patiënt die in Diemen verblijft positief voor het nieuwe coronavirus (COVID-19). Meer informatie: https://t.co/Bky2Ehwe2A #coronavirus #covid19 #corona https://t.co/mFJxJcCRbc
104	Een bron heeft ons laten weten dat in het Jeroen Bosch Ziekenhuis in Den Bosch mogelijk een aantal patiënten zich hebben gemeld met klachten dat lijkt op het Coronavirus. Het ziekenhuis heeft niet willen testen voor het virus. Het ziekenhuis is niet beschikbaar voor commentaar.
97	Ook enkele contacten van de Tilburgse patiënt testten positief voor COVID-19. Voor meer informatie: https://t.co/sahdK8RXdF #covid19 #covid #coronavirus #corona #virus https://t.co/8MGAcx5LGZ

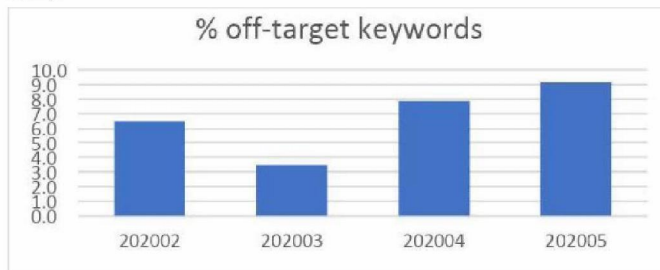
- 'Testmaterialen' became a more prominent topic over time, showing an increase from 11 % in February to 17.6% in May.
- 'Testbeleid' was a very frequent topic especially in March when the discussion centred upon the question why there was no large-scale testing for corona in the Netherlands. Two tweets in the top 5 tweets for this month actually relate to this:

freq	Tweet text
1438	"De Zwitserse farmaceut Roche is de reden dat Nederland niet massaal op het coronavirus kan testen. Het bedrijf is marktleider, kan basisvloeistof voor tests niet leveren maar wil het recept ervan niet vrijgeven. https://t.co/GEYatpr1jU "
910	De Zwitserse farmaceut Roche is een belangrijke oorzaak dat Nederland niet massaal op het coronavirus kan testen. Het bedrijf is marktleider en wil de laboratoria niet helpen door de receptuur van een essentiële vloeistof beschikbaar te stellen\n https://t.co/sep06vDaCy

- 'Testcapaciteit' was found to be one of the smaller topics, which reached a peak in April (4.2%).

In the analysis of the data we happened upon an unexplained difference in the coverage that we achieved: the coverage for the months April and May fell around 11 % behind that obtained for February and March. Upon closer investigation we found that part of the explanation was to be found

in the number of tweets that were actually off-target (i.e. not about corona testing). They were included in the dataset as a result of the collection method used: searching for tweets by means of the keyword 'test' inadvertently led to the inclusion of tweets where 'test' was matched in words like 'testimonium', 'ruimtestation', 'sterftestatistieken', etc. Of course a certain amount of noise is to be expected, only in this specific case we found that while in the February set the proportion of off target tweets was around 6.5 % and in March 3.4 %, in the months April and May this was up to 7.9 and 9.2% (the actual percentages are a bit higher for each of the months, for reasons we won't go into here).



A more detailed analysis showed that the off-target keywords could be categorized into 7 main classes (6 main class and 1 rest class – here referred to as 'anders'):

	202002 (%)	202003 (%)	202004 (%)	202005 (%)	All (%)
anders	0.3	0.4	0.9	1.1	0.7
attest	0.7	4.9	2.8	2.5	3.3
-est	7.6	9.0	4.2	2.8	5.2
protest	78.7	43.2	67.6	75.5	63.3
protestant	0.5	4.3	4.4	1.9	3.6
-test	7.4	23.5	6.9	5.3	11.1
testament	0.7	3.5	2.2	0.3	2.0
	100.0	100.0	100.0	100.0	100.0

The 'protest' class accounts for some over 63% of the off-target keywords (and tweets).

Although the proportion of 'protest' tweets is also quite substantial in February, most protest tweets in this period are about protests in Hong Kong and Iran, the climate crisis and more specifically in the Dutch context around the nitrogen crisis, the farmers' protests, and such.

In April and May, the orientation of the protest tweets is different: the main focus appears to be on criticizing corona policy (lockdown and other measures taken). There is an increasing number of tweets suggesting that corona is ploy by the government/government leaders (in Europe and at home) and a threat to democracy. For example,

- "Hoe ""onze""overheid de ""Corona-maatregelen"" MISBRUIKT om elke vorm van protest, hoe ""vriendelijk"" ook de kop in te drukken. Walgelijk @MarkRutte @Thierrybaudet @VVD <https://t.co/a5yMOLQiSC>"
- #Macron vindt die #lockdown van #Frankrijk (formeel wegens het #coronavirus) wel lekker. Niet meer die vervelende #gelehesjes elke week. Geen boerenprotesten meer. Geen protesten tegen verhoging van de pensioenleeftijd. Heerlijk rustig! <https://t.co/6uKFVelmh4>
- Maar oeverloos werd herhaald dat ze totaal voorbereid waren mocht er een uitbraak komen. Ondertussen zag Rutte in dat corona een machtig wapen zou kunnen zijn om de EU en Globalistische agenda door te voeren en protesten dmv 1,5 meter in de kiem te smoren. Heil Rutte ! <https://t.co/rnBXQrkWMO>
- #alleensamen gaan we versneld naar een politiestaat waar de angst regeert. EU Dictatuur en globalisme is het ultieme doel.\nRutte heeft nauwe banden met dictator Macron die protesten van zijn volk gewelddadig onderdrukte.\nDe coronalockdown was de 'natte droom'redding voor Macron.

- #CoronaApp is allemaal onderdeel van de agenda-21. Nog even en de politici hebben de milieu doelen, minder protesten en meer migratie voor elkaar #5g

Frequently asked questions on Twitter

Using [Relevancer](#) we clustered tweets from May 2020, divided into weekly bins, that contained a question mark ("?"). This selection contains tweets with questions of many different types. As any question can be asked in a variety of ways, it is not trivial to create a 'top X of questions asked' at any point. Relevancer, however, can be used to cluster tweets into groups of similarly phrased tweets. Indeed, when we clustered the "?" selection, about 75% of all clusters represented a coherent set of similar questions around the same topic. Manually we assigned labels to each topic, and we created weekly rankings of recurring topics in clusters. Ranking these questions in a top-4 for the five weekly bins of May 2020, we observe that there is a variance in the type of most frequently asked questions:



Questions are asked frequently about the consequences of the crisis for the economy; there is a continuous stream of questions on the interpretation of measurements, and critical questions on their logic. Some topics change in popularity: the topic of conspiracy theories dropped out of the top-4 after starting on number one in the beginning of May; the topic of adaptation to the new normal arrived in the top-4 in mid May and was the number one topic at the end of the month.

In more detail, we list topics occurring in more than one cluster (numbers are # of clusters):

May 1-3:

6	complot
3	kritiek_maatregelen
3	politiek
2	kritiek_testen
2	regio
2	depressieklachten
2	statistieken
2	coronaschulden

May 4-10:

11	onduidelijkheid_maatregelen
8	kritiek_maatregelen
6	economie
6	complot
5	vooruitblikken
5	compliance
5	ander_land
4	seks
4	statistieken
4	aanpassingen
3	kritiek_regering
3	geneesmiddel
2	religie

May 11-17:

20	economie
10	complot
10	kritiek_regering
6	vooruitblik
6	gezondheid
4	ontwikkelingshulp
3	onduidelijkheid_maatregelen
3	aanpassingen
3	geneesmiddelen
2	onlogica_maatregelen
2	racisme
2	vakantie
2	entertainment
2	statistieken
2	informatie_maatregelen
2	verpleeghuizen
2	misinformatie

May 18-24:

8	kritiek_regering
6	aanpassen
7	economie
5	slachthuizen
5	kritiek_maatregelen
5	vooruitblik
5	complot
4	vakantie
3	tweede_golf
3	waar_is_irma?
3	kinderen
2	ander_land
2	corona-app
2	compliance
2	ander_land
2	nertsfarms

May 25-31:

12	aanpassingen
12	economie
9	kritiek_maatregelen
6	onduidelijkheid_maatregelen
5	slachthuizen
4	corona-wet
3	nertsenfarms
3	complot
3	statistieken
2	regio
2	psychologische klachten
2	corona-app
2	humor

Other more constant topics aside from economics and critical questions on measurements are questions on statistics, politics (including critical questions on the government's actions), medical aspects of treating corona patients, corona outbreaks in mink farms and meat factories, holiday planning, the government's corona app, and fears of a second wave.