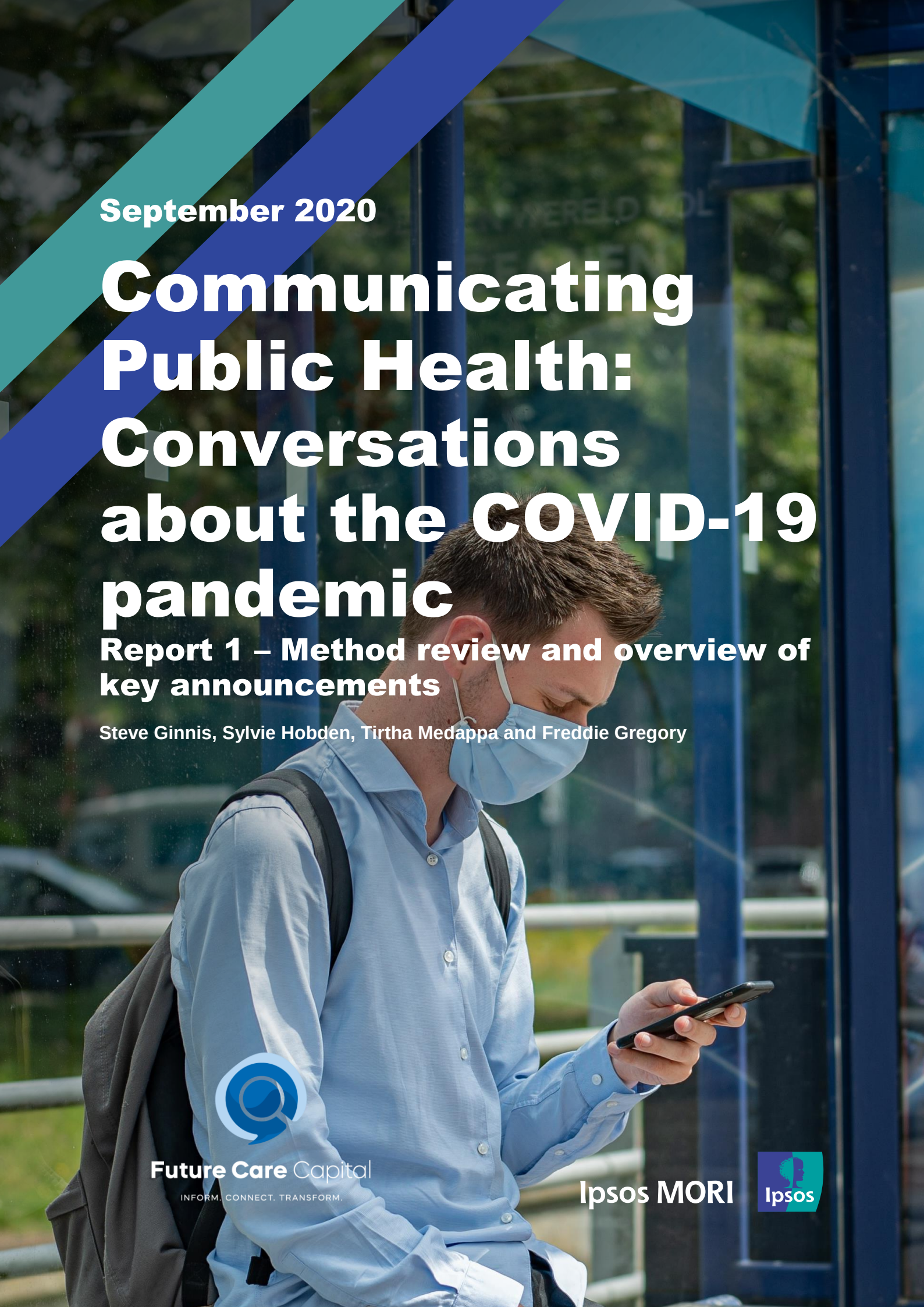


September 2020

Communicating Public Health: Conversations about the COVID-19 pandemic

Report 1 – Method review and overview of key announcements

Steve Ginnis, Sylvie Hobden, Tirtha Medappa and Freddie Gregory

Future Care Capital
INFORM. CONNECT. TRANSFORM.

Ipsos MORI

Contents

1 Foreword 4
2 Introduction and key findings 5
2.1 Background and objectives 5
2.2 Overview of approach 5
2.3 Key findings from Report 1 6
3 Method design and key considerations for interpretation 8
3.1 Overview of iterative design 8
3.2 Representativeness 10
3.3 Measurement 12
3.4 Ethics and care for social media users 14
3.5 Value of social media data and data quality 14
3.6 Conclusion 15
4 Key issues present on social media during the pandemic 16
4.1 Overview of key peaks in social media conversation 16
4.2 Sustained interest over time 17
4.3 Engagement with key announcements 18
4.4 Role of mainstream media 19
4.5 Conclusion 21
Appendix 22

1 Foreword

With our growing reliance upon social media platforms to share information, express ourselves and facilitate public dialogue, the debate about whether social media constitutes a force for good or something harmful to society continues to rage.

On the one hand, social media has transformed the way key messages are conveyed by the governments, businesses and civil society - enabling real-time dissemination of the latest news, advice and guidance as well as broad-ranging conversations and connections. There is, however, unease about some of the intensely personal effects of social media usage, mounting concern about the scope for platforms to perpetuate mis- and disinformation, as well as real-world implications which flow from their tendency to limit the voices we hear to certain groups or 'echo chambers'.

These issues are, by now, well-rehearsed and we make no judgement here about the relative merits of social media as a data-driven technology which has gained traction over a period of 15 years. Instead, our work with Ipsos MORI makes use of social media data to explore some of the things it can (and cannot) tell us about the unprecedented times we continue to experience as a result of the COVID-19 pandemic and, in particular, we take conversations about health and social care as our focus to evaluate their potential to usefully inform policymaking and public health interventions.

In this first report in a series, we detail our approach – its pros and cons - and endeavour to distil the 'key' announcements which have influenced social media activity in the UK during the period analysed. We identify the policies which gained greater traction online than others. We also begin to illustrate the way in which critically important services, their users and their families as well as those dedicated to providing them, are more or less 'heard' in and above the social media maelstrom.

We are living in the midst of the first data-driven pandemic which, no matter the experience, is an intensely visceral one. Data is in some senses our guide, something which mediates our knowledge and understanding of the events in which we are embroiled, and not always one we welcome when it is relied upon to convey concerning news. However, it also underpins the public health interventions that are designed to keep us safe, together with discovery and deployment of the treatments that scientists continue to strive to discover and make safe for widespread use. Each of our reports is designed to inform those tasked with communicating important messages about health and care in challenging circumstances.

Annemarie Naylor, Director of Policy and Strategy

Future Care Capital

2 Introduction and key findings

2.1 Background and objectives

In April 2020, during the COVID-19 lockdown, almost half of the British public (47%) reported that they were spending more time than usual on social media[1]. It appeared that, due to the strict physical distancing measures, people were turning to social media to maintain connectivity. The implication of this finding is that, during the pandemic more than ever, social media data should provide a useful insight into the perceptions and experiences of a sizeable proportion of the public.

The research described in this paper aimed to harness this opportunity, in order to understand how social media users responded to COVID-19-related information disseminated by the Government, the mainstream media and other sources.

This insight is urgently needed. The World Health Organisation (WHO) has shared its concern that in fighting the COVID-19 pandemic, they must also combat the "infodemic"[2]. This overabundance of information can cause important messages to go unheard, whilst misinformation or speculation gains traction. Understanding the nature of the content social media users generated and engaged with, and the way in which they did so, can inform the way in which information is best shared in the future.

Ipsos MORI were commissioned by Future Care Capital (FCC) to explore the insight that could be generated from analysis of social media under four specific aims:

1. Understand what constitute the ‘key’ announcements made in relation to the pandemic influencing social media activity in the United Kingdom (UK);

2. Evaluate which public health messages have ‘performed’ better or worse than others amongst social media users;

3. Explore the key trends and timelines associated with a range of challenges and opportunities impacting health, care and allied professionals and their ‘resolution’, as told on social media;

4. Draw on social media data to better understand the implications of the pandemic in the UK amongst social media users relating to mental and physical health.

2.2 Overview of approach

The primary methodology used to explore these aims was social media analysis. The analysis was based on an initial dataset of 3,692,129 social media posts from between 01 February and 30 June 2020. The dataset included posts from social networks and forums together with comments on key press articles.

[1] Ipsos MORI. (2020). *Social media, gardening, books, bread and having sex: How Britons are whiling away coronavirus lockdown.* https://www.ipsos.com/ipsos-mori/en-uk/social-media-gardening-books-bread-and-having-sex-how-britons-are-whiling-away-coronavirus-lockdown

[2] Zarocostas, J. (2020). *How to fight an infodemic*. The Lancet, 395(10225), 676. https://doi.org/10.1016/S0140-6736(20)30461-X.

This social media dataset was analysed using a combination of automated machine-led analysis, manual coding, statistical modelling and qualitative investigation. Social media analysis was augmented by desk research, survey research and Google Trends search analytics.

A detailed overview of the approach is provided in Chapter 3 below, alongside an exploration of the key merits and limitations provided by social media when seeking to understand the experience and views of members of the public.

This report is the first of four outputs from the research. It provides an overview of the methodology, and a discussion of the key considerations for judging the utility of social media data. It also outlines the findings relating to Objective 1 – understanding key announcements on social media relating to the pandemic.

Further findings from the project can be found in reports 2-4:

- Report 2 – Performance of public health messages and the impact of emerging issues
- Report 3 – Discussions about and among health and social care professionals
- Report 4 – Understanding mental and physical health among social media users during lockdown Key findings from Report 1

2.2.1 Value of social media data

- The social media analysis conducted as part of this project is not representative of the wider UK general population; instead, it seeks to be reflective of those who posted publicly about the coronavirus pandemic on social media. It is therefore important to note that it over-represents the views of younger adults and those from more affluent backgrounds. The findings are also reflective of the varied nature of social media platforms, both in frequency of posts and in access to publicly available data. The data discussed in this report is naturally weighted towards Twitter content, and towards the earlier stages of the pandemic which witnessed the highest levels of relevant posts on social media.
- Despite its imperfections, social media data provides a rich insight into the daily lived experiences of a proportion of the public during the coronavirus pandemic, and an unrivalled real-time perspective of how events, government communications, media coverage and public opinion interlink during a time of crisis. Rather than being compared to a quantitative survey, the strengths of social media in the context of the experience of the coronavirus pandemic lie in its ability to capture in-the-moment experience and opinion, and to compare the relative importance and performance of key issues over time.
- When using social media data for research, analysts and policy makers should be aware of a number of inherent risks associated with data collection, and of the mechanisms by which to judge the quality of the data. This should include close scrutiny of the data collection process, and efforts to understand the context in which people are posting on social media about the topic of interest – either through desk research or primary research.

2.2.2 Key announcements

- Throughout the pandemic, the policy areas that generated the highest volumes of discussion on social media were those that affected the vast majority, if not all, of the general public. These included lockdown guidelines, testing and vaccines and the Government's recovery strategy.

- Conversations about policy areas that affected subsets of the population - including quarantine guidelines, shielding guidelines and social care – were less evident on social media. Together, these findings imply that social media is a poor tool for communicating with those who are most likely to be impacted by such policy announcements.

- The data presented here provides indicative evidence that there is, at the very least, a symbiotic relationship between social media content posted by mainstream media outlets and that posted by members of the general public. It seems likely that a high level of mainstream media activity has the potential to propel conversation about key issues at the expense of others which receive less attention. There is therefore a risk that some policy areas receive little critical evaluation and debate on social media compared to other policy areas. This has implications for awareness, understanding and discussion of, for example, social care in the UK.

- Two further characteristics were also evident within the formal announcements led by government and public health bodies. Firstly, that public health messages have to compete with, and can be derailed by, unplanned events. Secondly, that future public health messages will need to overcome a falling rate of discussion about coronavirus on social media if the announcements are to be widely heard.

3 Method design and key considerations for interpretation

Key findings

The social media analysis conducted as part of this project is not representative of the general population; instead, it seeks to be reflective of those who posted publicly about the coronavirus pandemic on social media.

The quality of social media data is hampered by availability of data (where access is determined by platforms), and by the pragmatic data collection processes when searching for relevant social media posts.

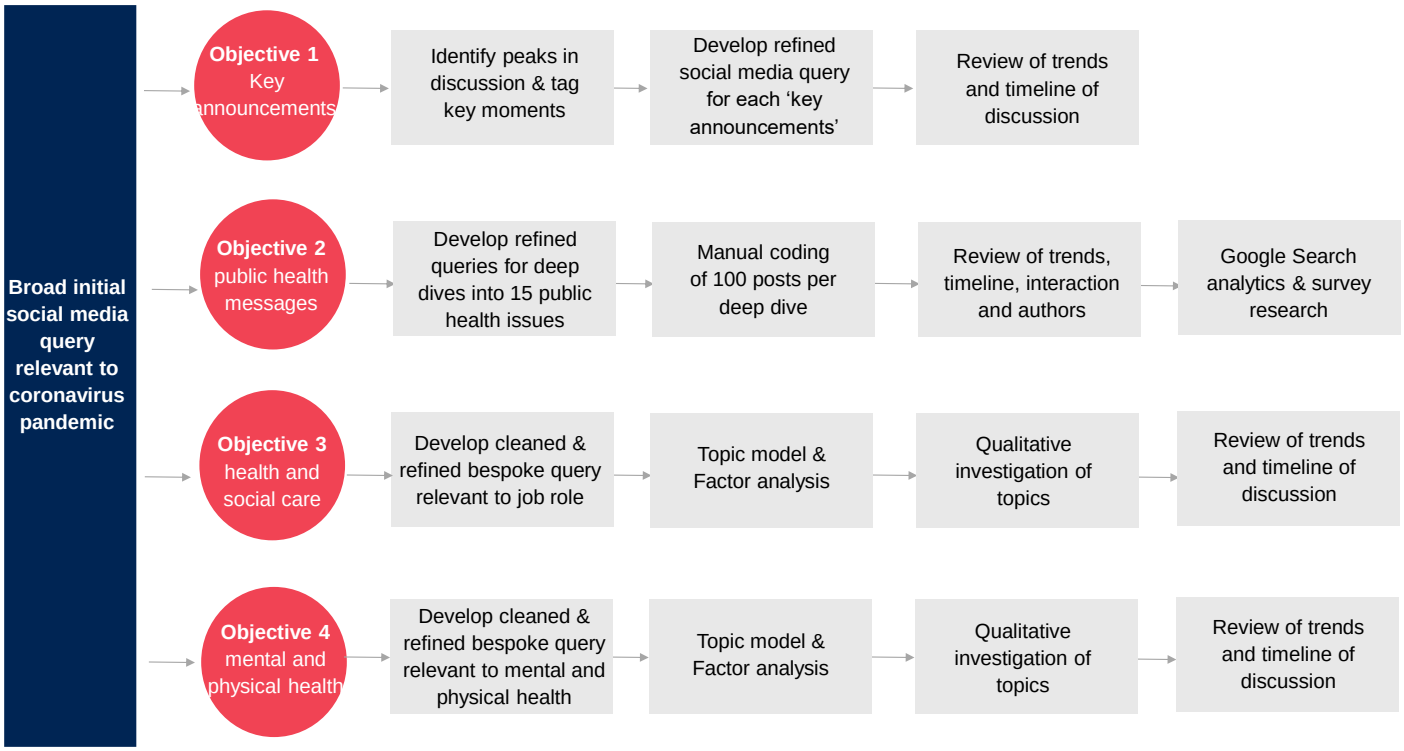
The characteristics of the data set are organic and context specific – they reflect natural skews in how and when users choose to post.

Despite its imperfections, and with due care and consideration, social media data has the potential to offer valuable insight in to the lived experiences of the public and health and social care professionals during the coronavirus pandemic.

3.1 Overview of iterative design

Four iterative pathways were developed for addressing each of the four objectives outlined above (see Chapter 2). An iterative approach allowed for data to be refined over time, and for deeper analysis to be conducted on the most relevant topics of interest. An overview of this approach is provided in Figure 3.1. Our approaches to addressing all four objectives were inspired by a broader initial social media search query that identified posts relevant to the coronavirus pandemic and associated announcements. Specific lines of inquiry were then followed within smaller samples of social media data.

Figure 3.1: Overview of methodology across four objectives



The project draws on three primary sources of data:

- **Social media data:** Publicly available social media data was collected using the Synthesio social listening platform – a software that aggregates social media data from multiple social media services. The key terms identified from the desk research were used to build an initial search query to extract social media posts mentioning COVID-19-related events. The search query sought to capture all relevant publicly available posts across social networks, forums, and news website comments. Additional parameters were set to only include posts in English language, and posted from within the UK. The result was an initial dataset of 3,692,129 social media posts from between 01 February and 30 June 2020.
- **Google Trends search analytics:** Using the capabilities provided by Google Trends, search activity related to each of the public health messages and issues of interest was analysed. This analysis included volume of relevant search activity over time, and assessment of the most common topics related to these search terms.
- **Survey of the general public:** A survey was conducted amongst a representative sample of UK adults, aged 16-75, to explore attitudes towards the government communications, and key issues relating to use of technology and data during the pandemic. The survey was conducted using Ipsos MORI's online Omnibus platform and received a total of 1,105 responses. Fieldwork took place 26-29 June 2020. Data was weighted by age, gender, region and work status.

Data sources used throughout this study

To aid interpretation of the data used throughout the study, each Figure displays a flag in the upper-right corner to identify its data source.

- A **blue** flag signifies that the data drawn from social media
- A **purple** flag signifies that the data is from the representative online survey
- An **orange** flag signifies that the data is drawn from Google search activity

There is currently much debate over the quality and use of social media data within research. In assessing the value of the data, and in interpreting the findings, it is important to consider potential errors of representativeness and errors of measurement in the context of this project.

3.2 Representativeness

One of the most important principles of social research is representativeness; that the attitudes that are studied reflect the views of a known group. Representativeness is not always important, but it is vital when the research aims to draw inferences about a wider population; for example, if the findings of this research were to be generalised to the wider population of the UK.

Traditional offline research ensures representativeness either by selecting research participants on the basis of them being reflective of society based on a range of factors, or by controlling for these factors after data collection. These factors often include age, gender, ethnicity and socioeconomic status.

It is difficult to apply these same methods to social media research. Challenges include: i) demographics of users can be impossible to obtain; ii) not all social media platforms give equal access to data; iii) social media data collection queries are likely to include some margin of error; iv) analysis tends to count the number of posts rather than the number of people; and v) the users and posts represented by any proportion of a given conversation online will be different (not all users post about all of the topics of interest).

For clarity, the social media analysis conducted as part of this project is not representative of the general population; instead, it seeks to be reflective of those who posted publicly about the coronavirus pandemic on social media.

3.2.1 Profile of users

We are able to draw on some survey data to help better understand the demographics of which members of the population are under and over-represented within the social media data collected for this report.

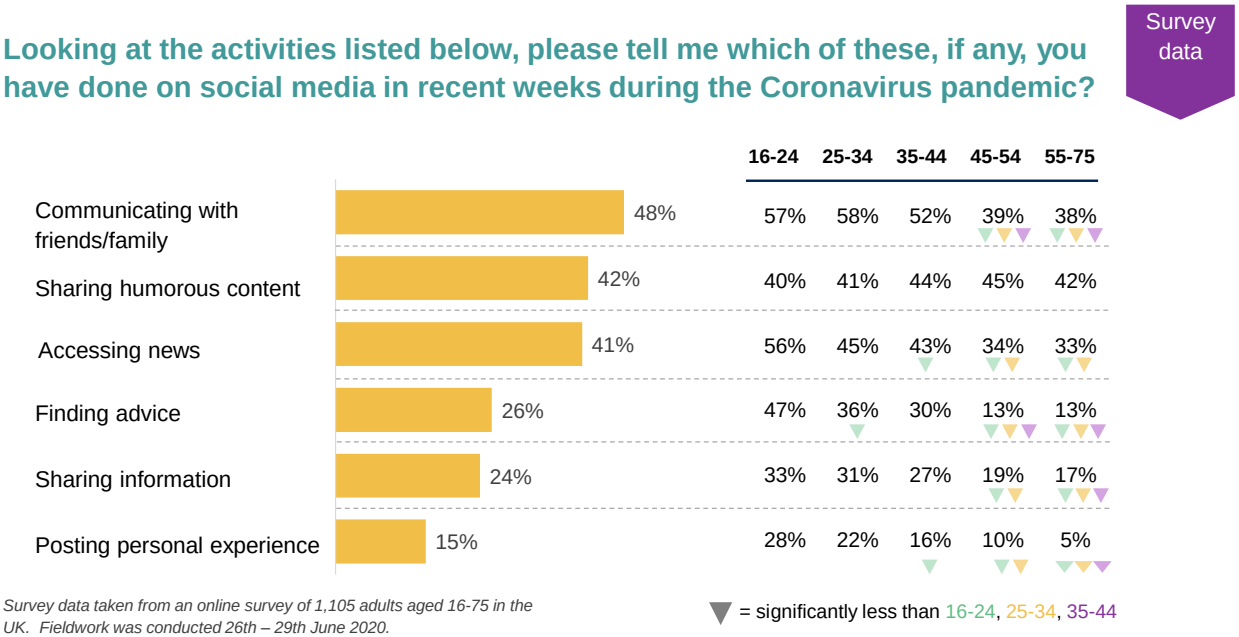
Firstly, although around 45 million people use social media in the UK, this only broadly equates to two thirds of the UK population.[3] Ipsos MORI's Tech Tracker captures variations by platform – although 60% of adults have a Facebook account only 24% use Twitter. Whilst the characteristics of users vary by platform, across the board, users are more likely to be younger adults and to be of

[3] Statista; UK: Active social media users 2020, Published by Joseph Johnson, Mar 3, 2020

higher social grade.[4] Conclusions from social media data are therefore more representative of younger, more affluent adults.

Secondly, social media data is context specific. Our survey of social media users during the pandemic shows that users were far more likely to be sharing humorous content and communicating with friends and family on social media than to be sharing information or posting about their own personal experiences. More broadly, social media was used to help access news and advice. Across all activities except for sharing humorous content, the use of social media is further skewed towards younger adults. Only a small proportion share their own personal experiences.

Figure 3.2: Social media activity relating to coronavirus



Survey data taken from an online survey of 1,105 adults aged 16-75 in the UK. Fieldwork was conducted 26th – 29th June 2020.

Finally, it is worth noting that Google search data and the online survey are likely to be representative only of the online population. Although Google represents a high market share of internet searches[5], the profile of users will, at best, only reflect those who use the internet. This similar limitation also holds true for the survey conducted online.

3.2.2 Profile of and access to data

It should also be noted that the data discussed in this report is weighted towards Twitter content, and towards the earlier stages of the pandemic.

As the most accessible source of social media data, it should not be surprising that Twitter posts constitute the highest source of content – representing 86% of data in the 'announcements' query , 79% of the data selected for objective 3, and 88% for objective 4. However, this is likely to represent a significant skew towards posts shared on Twitter; other platforms such as Facebook do not give

[4] For example, Twitter users jump to 38% of those aged 15-34 and 31% of those from ABC1 backgrounds, compared to 8% of those aged 55 and 15% of those from C2DE backgrounds. Ipsos MORI Tech Tracker Data Book, Q1 2020 https://www.ipsos.com/sites/default/files/ct/publication/documents/2020-04/tech_tracker_q1_2020.pdf

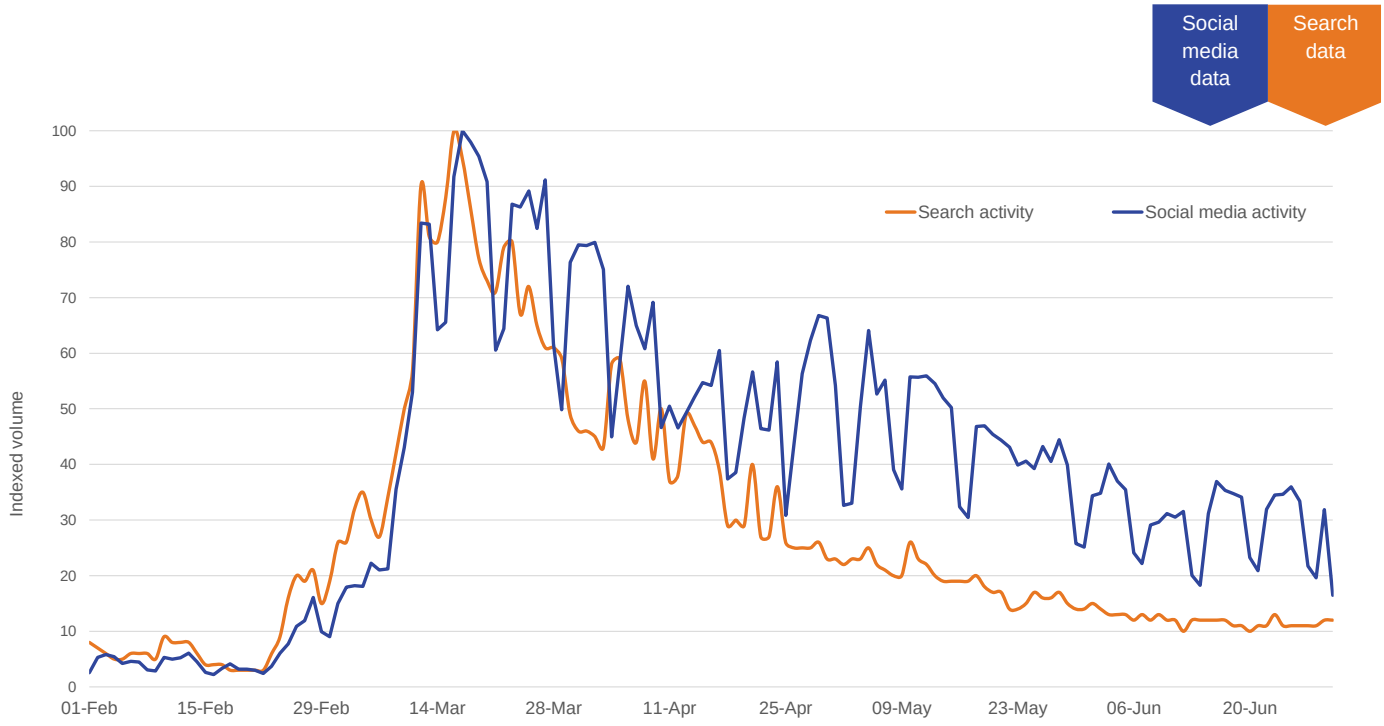
[5] Some estimates suggest as high as 93% in the UK: https://gs.statcounter.com/search-engine-market-share/all/united-kingdom

the same access to raw social media data for research. Data has not been weighted to seek to correct for this, as there is no accepted profile of relevant social media posts to weight the data to.

Furthermore, access to Twitter data is not unlimited. For the purpose of this project the Twitter Search API was used; this relies on building a Boolean search query to try to identify relevant social media posts. This provides a sample of all available data; yet it is difficult to calculate what proportion of data has been sampled as the total volume of posts relevant to a topic will be unknown. In theory access to the Twitter Firehose would generate access to an unbiased dataset; however in practice, analysts would still need to filter data to try to identify relevant posts.

The skew in conversation towards the start of the pandemic reflects the relative volume of content posted on social media over time. As shown in Figure 3.3 below, this general decline is supported by search data, which yields less searchers for coronavirus over time.

Figure 3.3: Social media posts and search queries relating to coronavirus over time



Base: Google searches of 'coronavirus', in the UK / Social media posts relating to coronavirus in the UK, between 01 Feb 2020 and 20 June 2020

Source: Ipsos MORI analysis of Google Trends/Synthesio data

3.3 Measurement

3.3.1 Data collection

A key component of errors of measurement is in the design of 'queries' used to collect data directly from a range of platforms. A 'query' is a search formula that uses keywords (which are not case sensitive) and Boolean operators (AND, OR, NOT, NEAR) combined with the intention of collecting only relevant social media data. Designing search queries is both an art and a science. A query that is too broad will collect too much irrelevant content whilst a query that is too restrictive will miss some relevant data. This project took a number of steps to try to improve the quality of the data collection process.

When developing queries for this research, desk research and iterative testing were utilised to inform the balance between false positives and false negatives. For example, the initial query relating to coronavirus events was informed by the Health Foundation's COVID-19 Policy Tracker.

The design of the query relating to objective 3 took a novel approach: the main body of the query was built around the statements which made it clear the person posting was a healthcare professional. For example, “as an NHS worker”, “I work in social care” and “I am a paramedic”. This data was supplemented with COVID-19-related posts from organisations that represent these groups. However, this approach also had limitations, as it was difficult to establish the context in which people belonged to these professions – for example paid or unpaid carers – where this was not specified by the user.

Only posts relevant to the UK were in scope for this project; however, not all posts contain location data. For clarity, the decision was taken to only include posts where location information was available; other studies may wish to take a more expansive approach.

A further data cleaning process was conducted in advance of the topic model and factor analysis, which would have been ineffective if the data contained too much irrelevant content. Here, machine learning algorithms were used to identify relevant posts. In order to train the algorithm, the research team manually coded a random selection of 500 social media posts as either broadly relevant or irrelevant to the challenges of health and care professionals. The algorithm was run on Neural Networks which recognises patterns and differences in the two groups of social media posts (relevant and irrelevant). Based on this analysis of pre-coded posts, the algorithm was able to then make predictions about how to classify data that had not been manually reviewed. For this project, the accuracy of this process was 72%, this was considered a high enough threshold to further enhance the relevancy of the data.[6]

Despite best efforts, the challenges of obtaining high quality social media data were evident when exploring the implications on mental and physical health (objective 4). Only having collected the data and begun analysis was it clear that some of the conversations relating to efforts to improve physical health during lockdown had been missed from the original Boolean search query. This is because social media comments were more informal in tone and often disconnected from health outcomes.[7]

3.3.2 Data analysis

Platforms such as Synthesio offer significant value in their ability to investigate social media data through the use of automated analytics; this includes metrics such as ‘interaction’ or automated ‘sentiment. However, it was important to supplement these metrics with bespoke analysis designed to meet the specific objectives of the research project.

For the purpose of evaluating the performance of public health messages (objective 3), manual coding of 100 randomly selected posts within each topic of interest provided an assessment of

[6] This process cleaned only irrelevant content, rather than seeking to identify bots. Future projects may wish to implement further cleaning to remove bots; however, the very presence of bots and the messages they promote is in itself a valid research phenomenon when seeking to understand the infodemic.

[7] For example, asking people to ‘get active’ without saying anything about ‘losing weight’ or ‘improve fitness’

emotion and humour, and the extent to which comments related to adherence, clarity of understanding, or support of the underlying proposition.

For the purpose of exploring the conversation relating to health and social care professionals (objective 3) and the conversation relating to mental and physical health (objective 4), bespoke statistical analysis was conducted to categorise and group the content of posts together. This included Topic Modelling[8], and Factor Analysis[9]. The use of bespoke models for each dataset provided a more granular level of understanding of the key themes within the data, which could then be examined over time.

3.4 Ethics and care for social media users

All research projects seeking to draw on social media data should consider how best to safeguard social media users, above and beyond the legal requirements of regulations such as the General Data Protection Regulation (GDPR).

Ipsos MORI have established a cautious approach to analysis of social media data, based on primary research with members of the public.[10] A number of key principles have been applied across the analysis conducted for this project: only publicly available data posted on social media platforms has been collected; this data has been processed in line with the terms and conditions of each platform; all raw data has been held securely and confidentially; where illustrative quotes have been used, these have been scrambled and anonymised to remove the risk of reidentification; where authors have been specified, these are either from 'blue tick' accounts or accounts with followers of over 1,000 people; finally, all identifiable data will deleted within three months of the project end.

3.5 Value of social media data and data quality

There is considerable value in using social media data to better understand the public experience of coronavirus, despite its imperfections. The strength of social media data isn't in claims of representativeness, but in the depth of insight provided in the moment by social media users.

Firstly, it provides a rich insight into the daily lived experiences of a proportion of the public, without risk of error in respondent recall. Without being prompted by a research question, social media users are open to share any experience, behaviour and opinion. The relative weight and nature of discussions on social media can also be telling, including which topics gained little attention or traction.

Secondly, across all three reports, social media data provides insight into how social media users responded immediately (and over time) to key events, announcements and media coverage. For example, we were able to see the impact of events, such as Dominic Cummings' trip to Durham, on

[8] Data was analysed using the Ipsos MORI in-house topic modelling platform, built in Python. This used natural language processing techniques to generate a list of terms and phrases that can bring meaning (for example, noun chunks; subject and object in the sentence; terms strongly associated with other terms). Term similarity was evaluated using a machine-learning algorithm focusing on meaning; words like "good" and "great", for example, were evaluated and classified as similar.

[9] A statistical factor analysis was then conducted to map relationships between subtopics and group them based on terms commonly occurring next to or near each other within a social media post

[10] Ipsos MORI's publications on this topic include: Practical Ethics in Social Media Research; #SocialEthics; Data Science Ethics Dialogue

the attitudes of social media users; and whether a conversation was organic or led by mainstream media coverage.

Thirdly, it offers an opportunity to assess how successful the Government was in delivering public health messages via social media – a key channel for communicating directly with the members of the public – and which other prominent individuals or organisations were successful in championing or critiquing these messages.

Finally, there is clear value in establishing a better understanding of the role of social media, in and of itself, during a time of crisis. For example, the data collected as part of the project demonstrates the wealth of advice and information available on social media throughout the pandemic (for members of the public, key workers and employers); it also demonstrates efforts by social media users to support each other with messages of positivity; and provides a valuable platform for understanding the interaction between those who hold different opinions or had different experiences during the pandemic.

3.6 Conclusion

No one source of data is perfect, all have strengths and limitations. Given the volume of data that can be captured by social media, it is tempting to judge it using the same criteria as one would evaluate the reliability of a quantitative survey; however, this is often unwise. A fairer assessment would perhaps be to judge the merits of the data from an ethnographic or qualitative perspective, to better understand the in-the-moment experiences of a population of interest, and how their journey evolves over time. This is of critical and unrivalled value for analysis of the experience of the coronavirus pandemic.

Yet it is also important not to overclaim the population represented by social media data. Social media users are not representative of the wider UK population as a whole; moreover, the precise population represented by any given topic of conversation on social media is context specific, and the data collection process is often imprecise. Where possible, research projects using social media should seek to better understand the characteristics of the social media users within their dataset, and seek to develop and refine high quality, relevant data samples through an iterative design. For example, this could include understanding the number of unique voices in the dataset, understanding limitations in access to data, and commissioning primary research to better understand which users post on a particular topic and in what format.

Social media not only a offers window in to the attitudes and opinions of wider populations, but it is also a phenomenon worthy of further investigation in its own right. The fight to win the ‘infodemic’ is still ongoing, with the next battle likely to be against attempts to discredit vaccination for COVID-19. Understanding the dynamics of social media, users, events and news organisations will be a crucial tool for better health outcomes.

4 Key issues present on social media during the pandemic

The first objective of the research was to understand, *based on social media activity*, what constitute the key announcements related to COVID-19.

To achieve a bottom-up understanding of the conversations taking place on social media, a broad search query was written to identify any social media posts relating to key announcements from the coronavirus pandemic between March and June 2020. Over 3.69 million social media posts were identified and subsequently collected.

Four lines of inquiry were considered to identify key announcements: i) peaks in volume of discussion; ii) sustained interest over time; iii) engagement; iv) relationship with mainstream media.

Key findings

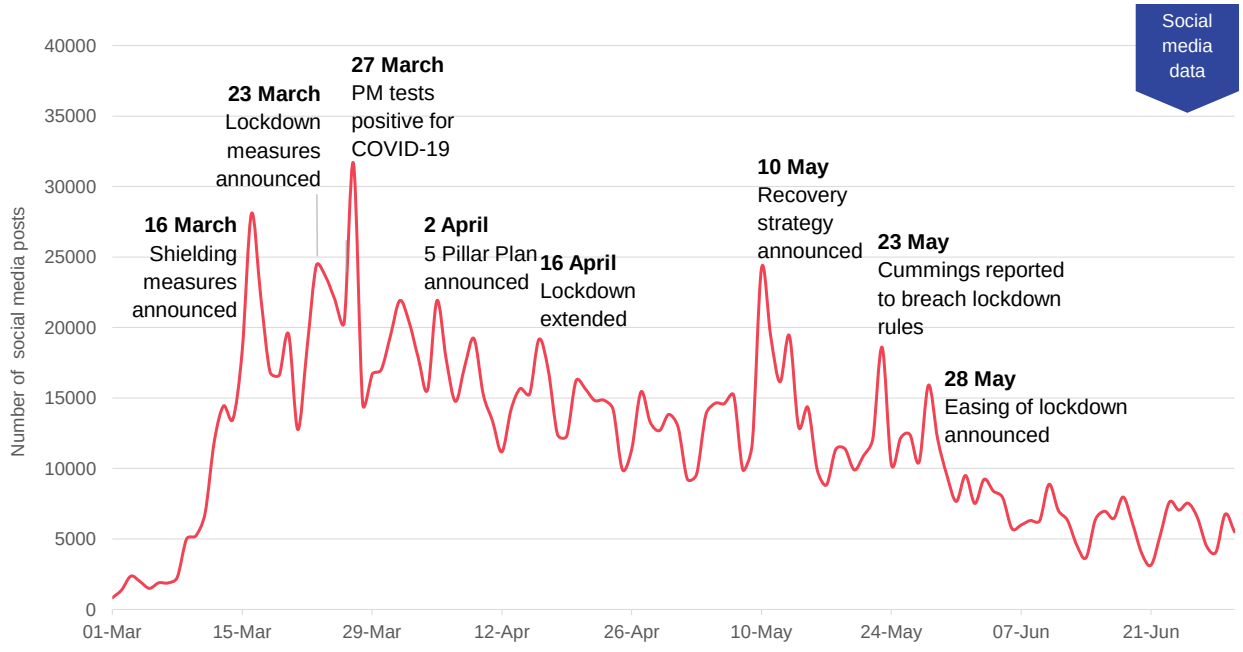
- Throughout the pandemic, the topics discussed in the highest volumes on social media were those that affected the vast majority of the general public. These included lockdown guidelines, testing and vaccines and the Government's recovery strategy.
- However, there was less interest in announcements relating to policy areas that affected subsets of the population - including quarantine guidelines, shielding guidelines and social care. Together, these findings imply that government should not seek to rely on social media as a communication tool on these issues; and that they received little critical evaluation and debate compared to other policy areas.
- The data presented here provides indicative evidence that there is, at the very least, a symbiotic relationship between social media content posted by mainstream media outlets and that posted by members of the general public. It seems likely that a high level of mainstream media activity has the potential to propel conversation about key issues at the expense of others which receive less attention.

4.1 Overview of key peaks in social media conversation

Figure 4.1 shows how the volume of these 3.69 million social media posts was distributed across the first months of the pandemic. The chart reveals a series of peaks, where the volume of social media posts related to COVID-19 increases quickly, before falling again. As such, the peaks in social media content capture spikes of immediate interest in a topic, rather than its sustained interest over time.

By reviewing the timepoints at which these peaks occurred, in combination with the findings of desk research, it was possible to identify which real-world announcements drove the most interest on social media.

Figure 4.1: Volume of COVID-19-related social media posts over time (overall)



Base: Synthesio data between 01/03/2020 and 30/06/2020
Source: Ipsos MORI analysis of Synthesio data

It is notable that the announcements which prompted peaks in activity are primarily those that affected the vast majority of the general public. Examples include the introduction of lockdown measures and the introduction of the recovery strategy.

There were also several peaks of activity related to individual personalities. For example, reports that the Prime Minister's advisor, Dominic Cummings, had broken lockdown restrictions prompted a peak in social media activity, as did the news that Boris Johnson had tested positive for COVID-19. To this extent, key unplanned events sparked just as much interest and discussion as planned government communications.

However, despite the breadth of events that drove the public to post on social media, a range of events are conspicuous by their absence. Analysis of volume of posts alone reveals little evidence that social media users engaged with announcements that only affected subsets of the general public. For example, there are no peaks associated with the flagship announcements relating to social care policy, quarantine guidelines or (other than its initial introduction) shielding.

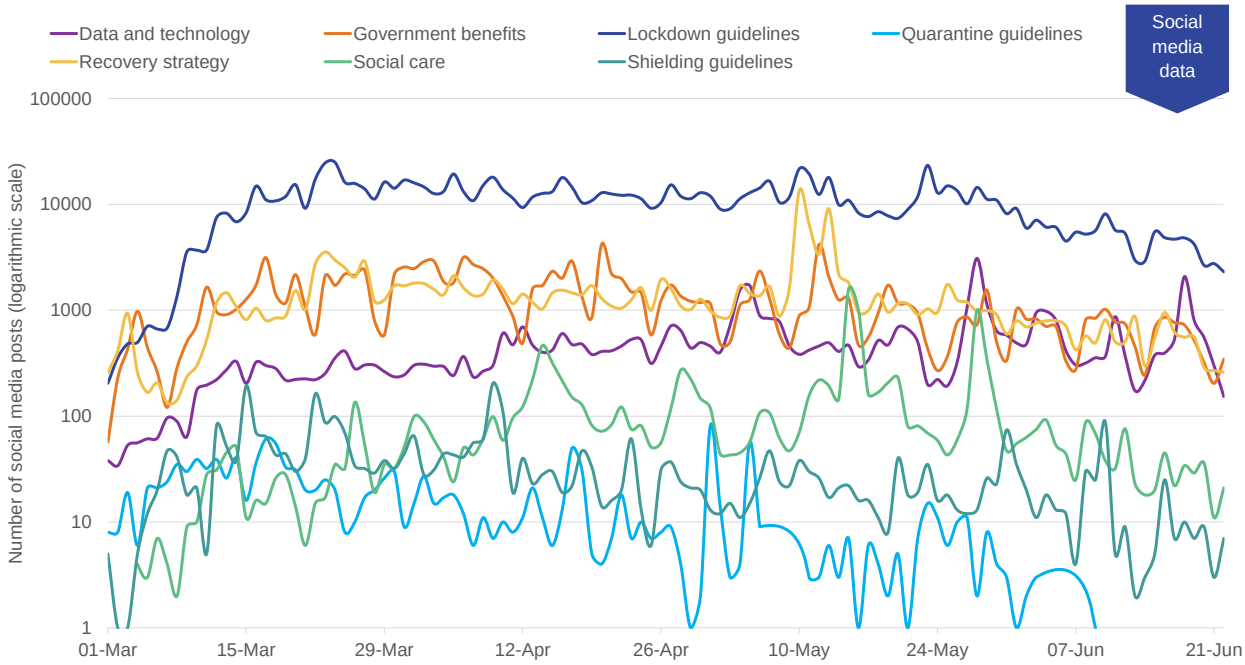
4.2 Sustained interest over time

Having established the key peaks in conversation, it is also important to consider which announcements sustained interest on social media over time. To achieve this, the overall conversation about COVID-19 that is presented in Figure 4.1 was split into separate areas of key policy announcements, derived from qualitative analysis of the Health Foundation's COVID-19 Policy Tracker.

Figure 4.2 presents the trajectory of conversation relating to these policy areas over time. The chart uses a logarithmic scale to enable differentiation between the less-discussed policy areas.

As is suggested by the overall data (Figure 4.1), the policy areas discussed in the highest volumes across the period related to those announcements which affected the general public (e.g. lockdown guidelines, testing and vaccines, recovery strategy). In contrast, announcements that related to policy areas that affected subsets of the population, including quarantine guidelines, shielding guidelines and social care, received comparatively little attention on social media.

Figure 4.2: Volume of COVID-19-related social media posts over time (by topic)



Base: Synthesio data between 01/03/2020 and 30/06/2020
Source: Ipsos MORI analysis of Synthesio data

It is perhaps unsurprising that these less-universal topics were discussed in relatively small volumes; after all, they have a direct impact on a smaller proportion of the population. However, the skew against these topics is likely exacerbated by the profile of social media users.

As discussed earlier in this report, those who use Twitter (from which most of the dataset originates) tend to be younger, better educated and more affluent than average. In contrast, those in receipt of either social care or shielding advice are likely to be older and less affluent than average. It would be reasonable to conclude that those affected by social care or shielding are less likely to be active on social media.

Together, these findings imply that, firstly, social media is a poor communication tool to engage those individuals who are most likely to be directly affected by key announcements relating to shielding and social care. Secondly, that the flagship policy announcements relating to these topics received less attention and scrutiny by other users of social media – including the interest of politicians, health groups, and private journalist accounts posting on social media.

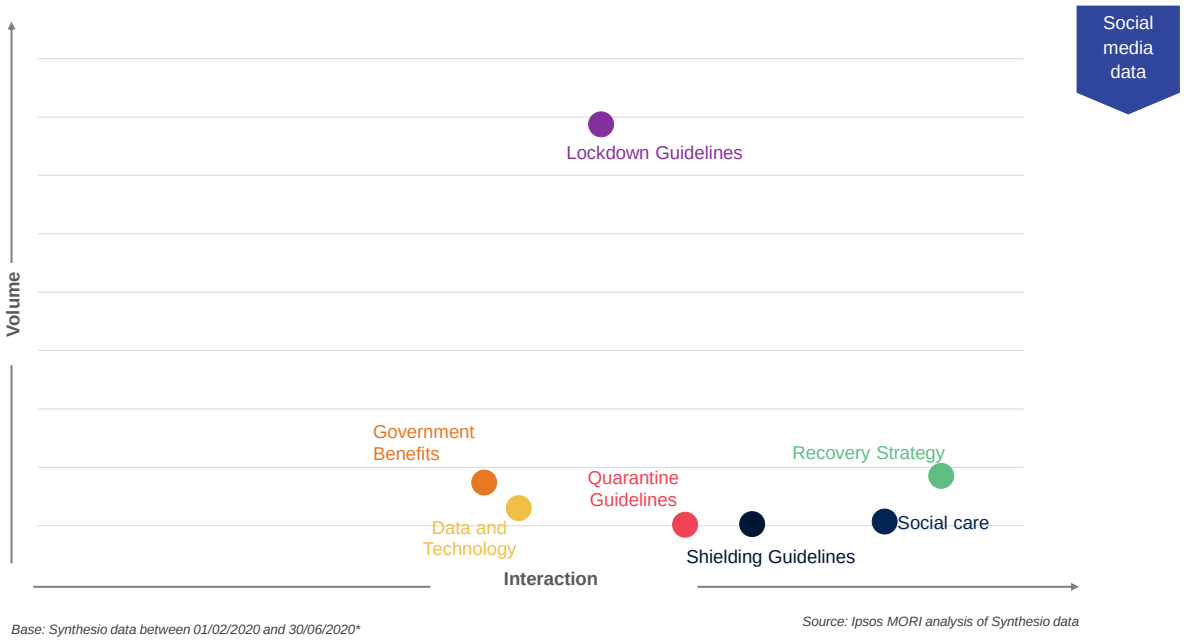
4.3 Engagement with key announcements

In addition to evaluating volume of posts, it is also important to consider the level of 'engagement' on social media relating to key announcements. The Government Digital Service suggests that engagement is a more meaningful metric than volume, as it allows an assessment of whether

people are actively consuming content. Engagement can be measured based on the number of interactions (e.g. 'likes', reposts, comments) that each post receives.

Figure 4.3 plots the number of average interactions against volume for each key policy area. It is notable that the topics which prompt the highest average interaction across the period are those which prompted some of the lowest volumes of related posts. This indicates that, although posts on these topics occurred less frequently, when they did, social media users were more likely to actively engage with them.

Figure 4.3: Average interaction with social media posts over time (by topic)



*Base: Synthesio data between 01/02/2020 and 30/06/2020**

Source: Ipsos MORI analysis of Synthesio data

Statistical analysis also suggests there is a strong correlation between the level of interaction posts receive and the sentiment of social media posts ($r(11,947)=.770$, $p<0.01$). Posts which have a negative sentiment, are more likely to prompt high levels of interaction from social media users.

Taken together these findings suggest that the high level of interaction with posts relating to social care policy and the Government's recovery strategy are driven by content that is negative or controversial in nature. From a methodological perspective, it also highlights the importance of analysing a range of metrics in order to obtain a comprehensive understanding of topics that are discussed in low volume.

4.4 Role of mainstream media

A striking feature of the social media data analysed was the volume of COVID-19-related content shared by mainstream media outlets. Between February and June, posts from mainstream media accounts, comprised between 5% and 10% of social media posts identified as being COVID-19-related across the time period.

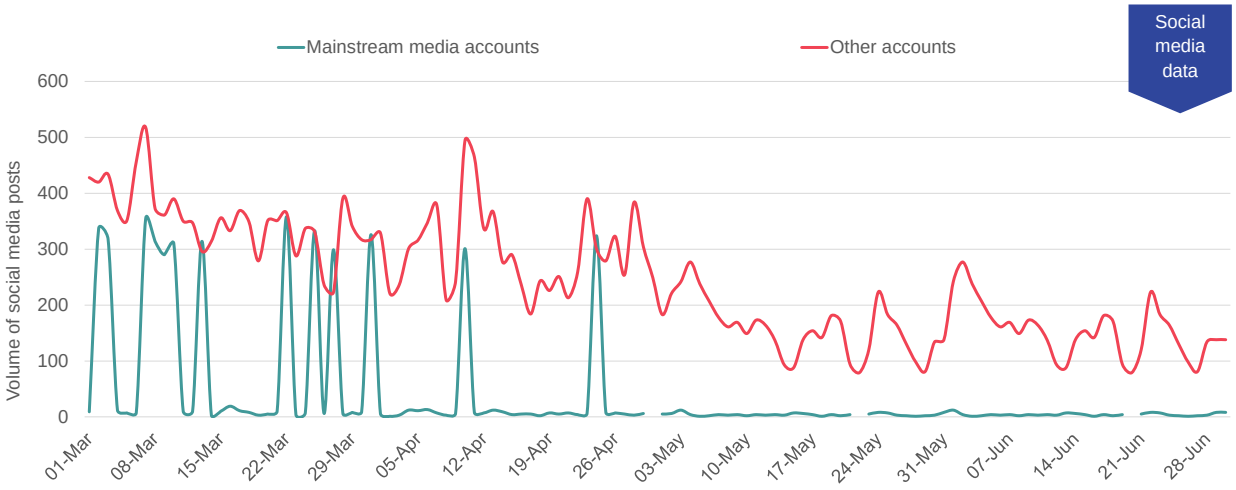
Given this finding, the social media analysis described thus far excluded content from mainstream media outlets. This decision was taken to ensure the analysed data reflected public experience and was not skewed towards mainstream media content.

Nonetheless, given the proliferation of content from mainstream media accounts, it is important to understand how their activity – in terms of the topics these accounts post about - is related to the activity of wider social media users.

To explore this question, the correlation between the volume of posts by mainstream media accounts and the volume of posts by other accounts, was assessed. Two lines of enquiry were explored: first posts that discussed individuals who had held a prominent role in delivering key government announcements;[11] and posts discussing wider issues of social care (which were not apparent in the overall data).[12]

Figure 4.4 presents the volume of posts that discussed prominent government officials, broken down by mainstream media and other accounts. Approximately 10% of posts on this topic originated from mainstream media accounts. Where the data indicates a peak in mainstream media activity on the topic, there is often a co-occurring increase in content from other accounts. Indeed, statistical analysis reveals a weak-to-moderate correlation between the volumes of content posted by mainstream media and other accounts on this topic ($r_s(126) = .32$, $p < .001$).[13]

Figure 4.4: Volume of posts discussing prominent government officials by account type



Base: Synthesio data (38,652 posts) between 01/03/2020 and 30/06/2020
Source: Ipsos MORI analysis of Synthesio data

Figure 4.5 presents the volume of posts relating to social care over time, broken down by mainstream media and other accounts. Although approximately 10% of posts relating to social care originated from accounts belonging to the mainstream media, the vast majority of these posts were in relation to the number of deaths in care homes. This is demonstrated by the peak in activity on 28 April, when deaths in care homes and the community were included in the COVID-19 related death figures for the first time. Otherwise, throughout the course of the pandemic, relatively few posts relating to social care were shared by mainstream media accounts. It is also worth noting that discussion of this issue was not sustained; and moreover, that the combined peak on and around 28 April still represents a relatively small share of voice in social media conversation compared to the spikes achieved on other topics presented at Figure 4.1.

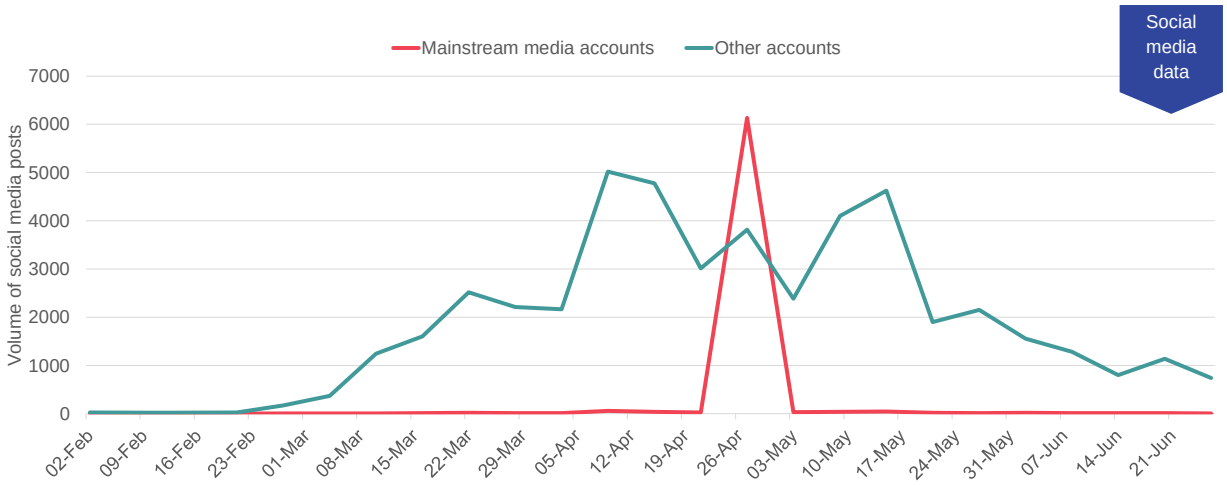
[11] For example, Chris Whitty, Rishi Sunak, Gavin Williamson or Simon Stevens

[12] For the purpose of this task, the social media query was expanded to include any references to social care and 'deaths'.

[13] The spikes in media coverage cover a range of topics, from early anticipation of announcement as UK makes 'extensive preparations' to move to next phase and personal pleas to 'stay at home', to discussion of Boris Johnson falling ill and then returning to work after recovery.

It could be hypothesised that the relatively low volume of conversation about social care (as presented earlier in this report see Figure 4.1) is in part due to the low coverage of the topic by the mainstream media. This in turn, may reflect the extent to which social care featured within the Downing Street Press Conference.

Figure 4.5: Volume of posts relating to social care by account type



Base: Synthesio data (50,000 posts) between 01/02/2020 and 30/06/2020
Source: Ipsos MORI analysis of Synthesio data

4.5 Conclusion

When considering this data, it is important to note that correlation does not imply causation. However, the data presented here provides indicative evidence that there is, at the very least, a symbiotic relationship between social media content posted by mainstream media outlets and that posted by members of the general public. It seems likely that a high level of mainstream media activity has the potential to propel conversation about events. Conversely, where mainstream media outlets post little on a topic, the topic is less likely to be posted about elsewhere. This is particularly evident in the case of policies relating to social care, which receive comparatively little attention on social media.

Further research would be required to ascertain the precise nature of the relationship between posts from mainstream media accounts and posts from the general public. This would benefit from inclusion of wider media monitoring data to compare coverage on other platforms such as radio, TV and print.

The relationship with news is also entwined with the nature of social media as an event driven platform. Spikes in discussion of Boris Johnson falling ill, and Dominic Cummings trip to Durham demonstrate that formal public health announcements from government have to compete with, and can be derailed by, immediate surrounding events.

The research also suggests that government and health organisations face a further challenge in being able to sustain interest over time. With the exception of the announcement of the recovery strategy on 10 May, interest and discussion of key announcements on social media broadly declines over time – this follows a broader pattern in falling rates of discussion about coronavirus more generally. Further public health communication strategy needs to succeed in bucking this trend if announcements are to be noticed widely.

Appendix

Bespoke social media search queries were developed to help track conversation relating to announcements from seven key areas of policy. A further two queries were developed to explore the role of mainstream media in driving social media conversation. The queries are provided below.

Data and technology

("central data store" OR "symptom checker" OR "contact tracing" OR "test and trace" OR (app NEAR ("isle of wight" OR iow OR nhs? OR apple OR google OR pivotal)))

AND

((wuhanvirus* OR coronavirus* OR novelcorona* OR 2019?nCoV OR "2019 ncov" OR 2019ncov OR covid19* OR "covid 19" OR "covid?19" OR covid) OR ((wuhan OR corona OR "cv-19" OR cv19) NEAR (virus OR disease)))

Government benefits

("business rates" OR "emergency response fund" OR "furlough*" OR "hardship fund" OR ("self*employed" NEAR "cash grant") OR "sick pay" OR "universal credit" OR "tax credits" OR (spring NEAR/2 budget) OR "job retention scheme" OR "business interruption loan" OR "emergency measures bill" OR "response fund")

AND

((wuhanvirus* OR coronavirus* OR novelcorona* OR 2019?nCoV OR "2019 ncov" OR 2019ncov OR covid19* OR "covid 19" OR "covid?19" OR covid) OR ((wuhan OR corona OR "cv-19" OR cv19) NEAR (virus OR disease)))

Lockdown guidelines

(lockdown OR (exercise NEAR (one OR once OR unlimited)) OR "work from home" OR ((essential OR necessary OR "non?essential") NEAR (travel OR contact)) OR "stay at home" OR ((café* OR pub? OR gym? OR restaurant* OR bar? OR shop* OR nighclub* OR theatre* OR cinema* OR "leisure centre*") NEAR (shut OR close)) OR ((autis* OR "learning disability*") AND (lockdown OR discrimin* OR rules OR relax* OR guidance)))

AND

((wuhanvirus* OR coronavirus* OR novelcorona* OR 2019?nCoV OR "2019 ncov" OR 2019ncov OR covid19* OR "covid 19" OR "covid?19" OR covid) OR ((wuhan OR corona OR "cv-19" OR cv19) NEAR (virus OR disease)))

Quarantine guidelines

((temperature OR fever OR cough* OR symptom* OR sniff* OR breath* OR resipiritory) AND ("self?isolate" OR quarantine) AND ("14 days" OR "fourteen days" OR "seven days" OR "7 days"))

AND

((wuhanvirus* OR coronavirus* OR novelcorona* OR 2019?nCoV OR "2019 ncov" OR 2019ncov OR covid19* OR "covid 19" OR "covid?19" OR covid) OR ((wuhan OR corona OR "cv-19" OR cv19) NEAR (virus OR disease)))

Recovery strategy

("recovery strategy" OR "three*stage" OR "3?stage" OR "exercise outdoors" OR "outdoor sports" OR ("international travel" AND isolate) OR ("non?essential" AND (retail OR shops)) OR "plan to rebuild" OR "road*map" OR "re?opening" OR "restart* the economy" OR "back to work" OR "return to work" OR "avoid public transport" OR "stay alert" OR "ease lockdown" OR "control the virus" OR "partial relaxation" OR "save lives" OR ((action OR response OR battle) NEAR/1 plan))

AND

((wuhanvirus* OR coronavirus* OR novelcorona* OR 2019?nCoV OR "2019 ncov" OR 2019ncov OR covid19* OR "covid 19" OR "covid?19" OR covid) OR ((wuhan OR corona OR "cv-19" OR cv19) NEAR (virus OR disease)))

Social care

((("care home?" OR "social care" OR "care setting*" OR "care provider*" OR "care sector") NEAR ("support package" OR "task*force" OR PPE OR outbreak OR discharge OR funding OR staffing OR budget OR "protective ring" OR ringfenced OR "wellbeing package" OR 85259 OR helpline OR "action plan")) OR ((restrict OR only) NEAR "one care home") OR (test* NEAR (care NEAR/2 "workers or staff")))

AND

((wuhanvirus* OR coronavirus* OR novelcorona* OR 2019?nCoV OR "2019 ncov" OR 2019ncov OR covid19* OR "covid 19" OR "covid?19" OR covid) OR ((wuhan OR corona OR "cv-19" OR cv19) NEAR (virus OR disease)))

Shielding guidelines

(("70 or older" OR older OR elderly OR "pre?existing health condition*" OR "underlying health condition*" OR "chronic disease*" OR "vulnerable" OR "high risk" OR "organ transplant" OR cancer OR scid OR "sickle cell" OR immunosuppress* OR "heart disease" OR "respiritory conditions") NEAR (cruise* OR shield OR "stay at home" OR "stay indoors" OR "outside"))

AND

((wuhanvirus* OR coronavirus* OR novelcorona* OR 2019?nCoV OR "2019 ncov" OR 2019ncov OR covid19* OR "covid 19" OR "covid?19" OR covid) OR ((wuhan OR corona OR "cv-19" OR cv19) NEAR (virus OR disease)))

Prominent government officials

("matt hancock" OR "chris whitty" OR "chief medical officer" OR "rishi sunak" OR "boris johnson" OR "prime minister" OR "gavin williamson" OR "health secretary" OR "simon stevens" OR "head of the nhs" OR "john newton")

AND

((wuhanvirus* OR coronavirus* OR novelcorona* OR 2019?nCoV OR "2019 ncov" OR 2019ncov OR covid19* OR "covid 19" OR "covid?19" OR covid) OR ((wuhan OR corona OR "cv-19" OR cv19) NEAR (virus OR disease)))

Social care

("care home?" OR "social care*" OR "care setting*" OR "care provider*" OR "residential home" OR "care sector" OR "care workers")

AND

(coronavirus OR covid*)

For more information

3 Thomas More Square
London
E1W 1YW

t: +44 (0)20 3059 5000

www.ipsos-mori.com
http://twitter.com/IpsosMORI

About Ipsos MORI Public Affairs

Ipsos MORI Public Affairs works closely with national governments, local public services and the not-for-profit sector. Its c.200 research staff focus on public service and policy issues. Each has expertise in a particular part of the public sector, ensuring we have a detailed understanding of specific sectors and policy challenges. Combined with our methods and communications expertise, this helps ensure that our research makes a difference for decision makers and communities.

Ipsos MORI