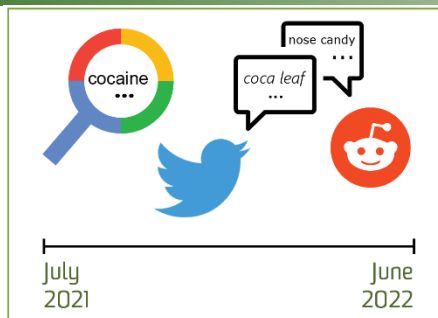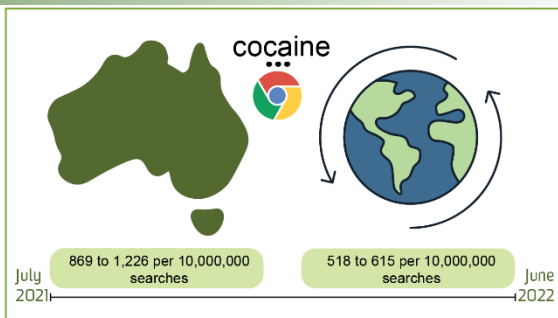# DRUG TRENDS

Trends in Google searches and social media discussions of cocaine, July 2021-June 2022: A pilot study

Qingyuan Linghu, Amy Peacock, Raimondo Bruno, Rachel Sutherland, Monica J. Barratt and Nicola Man

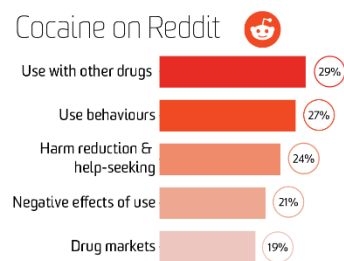*National Drug and Alcohol Research Centre*

University of New South Wales Sydney

We identified cocaine-related Google searches and posts on Twitter and drug-related subreddits from 1st July 2021 to 30th June 2022.

869 to 1,226 per 10,000,000 searches

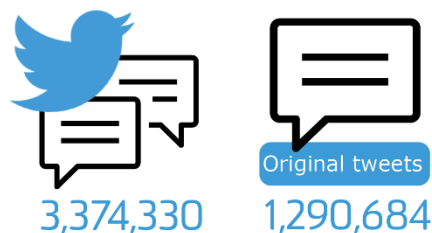518 to 615 per 10,000,000 searches

July 2021 — June 2022

The monthly Google search probability for cocaine was consistently higher in Australia compared with the worldwide searches from July 2021 to June 2022.
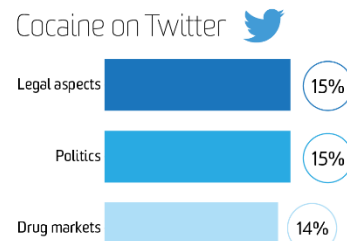

1,273,209 — r/cocaine

We have identified 1,273,209 posts about cocaine on Reddit from July 2021 to June 2022. They increased from July 2021 to April 2022, then plateaued and decreased until June 2022. Much of the increase over the 12-month period was from the largest cocaine-related subreddit "r/cocaine".

Cocaine on Reddit

- Use with other drugs — 29%
- Use behaviours — 27%
- Harm reduction & help-seeking — 24%
- Negative effects of use — 21%
- Drug markets — 19%

The most common topics discussed on drug-related subreddits in relation to cocaine.

3,374,330 — Original tweets 1,290,684

We have identified 3,374,330 tweets with reference to cocaine on Twitter, of which 1,290,684 were original tweets (i.e. excluding retweets). The monthly trend in cocaine-related original tweets increased from around December 2021 to April 2022, then plateaued and decreased in May and June 2022.

Cocaine on Twitter

- Legal aspects — 15%
- Politics — 15%
- Drug markets — 14%

The most common topics discussed on cocaine-related original tweets.

DRUG TRENDS

Key Findings

- We collected cocaine-related data on Google searches, as well as posts on Twitter and drug-related subreddits from 1st July 2021 to 30th June 2022.

- The monthly Google search probability for cocaine was consistently higher in Australia compared with the worldwide searches, suggesting cocaine-related searches were more popular in Australia compared with the rest of the world. It ranged from 869 to 1,226 per 10,000,000 searches in Australia and from 518 to 615 per 10,000,000 searches globally.

- Although fluctuating over time, the general trend suggested a decline in the probability of Australian cocaine-related searches over the period.

- We have identified 1,273,209 posts (including comments) about cocaine on drug-related subreddits from July 2021 to June 2022. Reddit posts increased from July 2021 to April 2022, then plateaued and decreased in May and June 2022. Much of the increase over this 12-month period was from the increase in number of posts on the largest cocaine-related subreddit, "r/cocaine".

- We have identified 3,374,330 tweets with reference to cocaine on Twitter from July 2021 to June 2022, of which 1,290,684 were original tweets (i.e. excluding retweets). The monthly trend in cocaine-related original tweets increased from around December 2021 to April 2022, then plateaued and decreased in May and June 2022.

- The most common topics discussed on drug-related subreddits in relation to cocaine were use with other drugs (29%), use behaviours (27%), harm reduction and help-seeking (24%), negative effects of use (21%), and drug markets (19%).

- The most common topics discussed in cocaine-related original tweets on Twitter were legal aspects (15%), politics (15%) and drug markets (14%).

- The discussions on drug-related subreddits are generally more positive than on Twitter. Law enforcement posts have the most negative sentiment score on both Reddit and Twitter.

DRUG TRENDS

Background

This bulletin reports on trends in Google searches and social media discussions of cocaine, as well as analysis of themes emergent in the latter. The current bulletin focuses on Google searches, and Reddit and Twitter discussions from 1st July 2021 to 30th June 2022.

This work was undertaken by Drug Trends to establish the feasibility of using Google search and social media data to inform illicit drug monitoring. Online data sources comprise a rich and relatively untapped source of information about illicit drug-related attitudes, markets, purchasing, use and behaviours. Online data can be accessed historically or in real time, are low cost to collate, and can provide important insights, particularly when studied alongside traditional epidemiological indicators. See here for our existing reporting on online data sources including trends in the availability and types of drugs sold on cryptomarkets.

The first step to collecting the cocaine-related online data was to determine a list of search terms for cocaine to identify cocaine-related activity in each online data source. The data were collected retrospectively using the search terms for cocaine for each of the three data sources. The social media posts (i.e. Reddit submissions and comments, and tweets) were then categorised into topics using keywords selected from a list of the 1000 most frequently occurring words. Unsupervised topic modelling was also performed to compute a list of potential keyword pairs to select as keyword pairs for topic categorisation of the posts. Sentiment analysis was then performed on each social media source, overall and by topic.

Note that we cannot guarantee exhaustive identification of all social media posts about cocaine. For example, the term '*coke*' may be used to refer to cocaine but it might also refer to the drink Coca Cola. The ambiguity or lack of specificity of some terms such as '*coke*' precludes their use in identifying cocaine-related activity on platforms less specifically orientated to cocaine such as in Google searches or Twitter even if these terms may identify cocaine-related searches or posts. Similarly, social media posts may not be exhaustively categorised into topics because of ambiguity of terms. The ambiguity of terms also means that some data may be wrongly classified into topics. As we have collected the social media data retrospectively, some of the Reddit posts and comments or tweets may be deleted or removed from the platform before they can be collected. Details on the methods and other possible caveats to interpretation are detailed in the Methods document.

DRUG TRENDS

It should also be noted that, for the purpose of this bulletin, social media data could not be geolocated to Australia and thus findings reported here for Reddit and Twitter reflect global trends. Options for study of social media data specific to Australia are being explored by the authors.

Monthly trend in the overall frequency of Google searches and social media mentions of cocaine is presented in this bulletin (section 1). In addition, we have presented the frequency of social media mentions of cocaine and the sentiment score by the identified topics, as well as corresponding monthly trends by topic (section 2).

Panel A. Terminology

- **Google search probability** is the probability that a short search session (e.g., a few consecutive searches) satisfy the search criteria (e.g., searches for cocaine terms) within a given location (i.e., Australia or worldwide) and time period (i.e., month). This is multiplied by 10,000,000 by the Google Extended Trends API for Health (GETH) for ease of presentation.

- **A tweet** is a message publicly shared on Twitter by a user[1], by which the user spreads information and interacts with others. Retweeting is the behaviour of a user forwarding another user's tweet, with or without additional comments. This bulletin reports on both original tweets (i.e., excluding retweets) and all tweets (i.e., including retweets), and we focus on the text of the tweet.

- **A subreddit** is a forum or group on Reddit, where the users talk about a certain topic which can generally be interpreted from the subreddit name, e.g., "r/cocaine" is the largest subreddit where users discuss cocaine.

- **A Reddit submission** is the content a reddit user shared to initiate a posting thread on a subreddit. We focus on the heading and body text of the submission.

- **A Reddit comment** is the content of a reply to a submission by reddit users in a subreddit, of which we also focus on the text of the comment.

- **A cocaine-related post** is the unified term referring to an instance of a tweet, a Reddit submission or a Reddit comment identified with a cocaine search term.

---

[1] A user in the context of online activity refers to someone who uses Google, Reddit or Twitter in this bulletin. When we refer to someone who uses drugs, we use a term such as "people who use drugs".

There are various approaches to collecting, collating, categorising and analysing the online data, and inherent challenges exist in these processes. For this reason, we have attempted to be as transparent as possible about our procedures. Further background and information regarding the methods are available for download. We welcome feedback and suggestions so that we can continue to improve utility of these data and our reporting on them drugtrends@unsw.edu.au.

Findings

1. The monthly trend in cocaine-related searches and social media mentions from 1st July 2021 to 30th June 2022
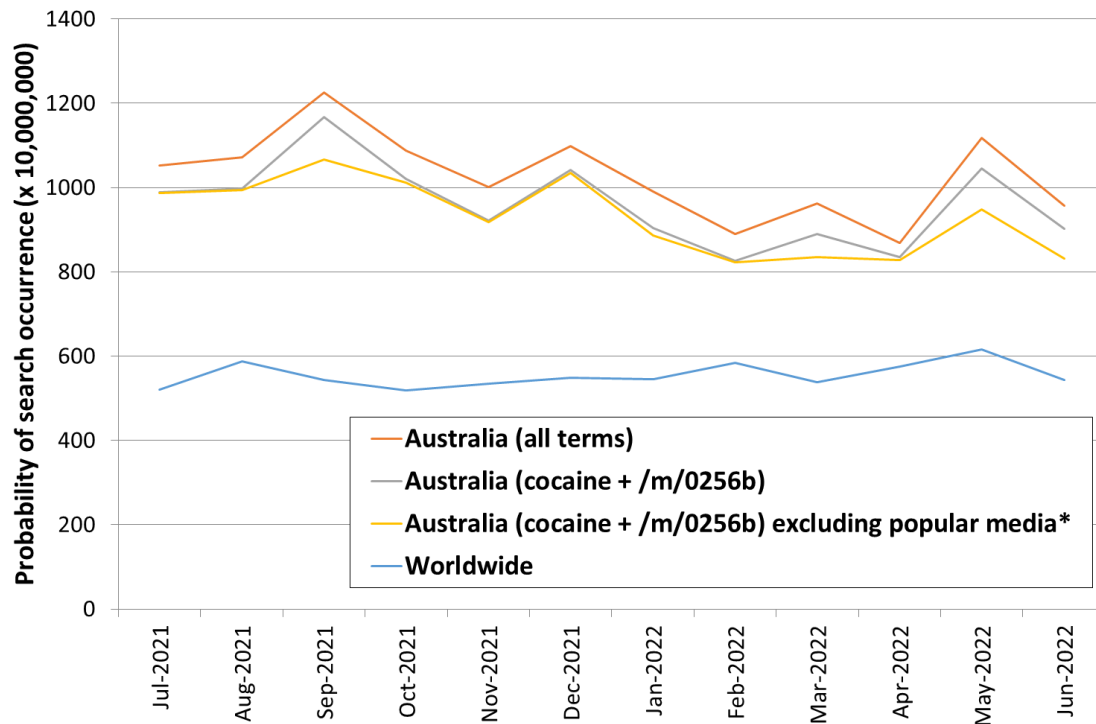
1.1. What is the monthly trend in cocaine-related Google searches?

Figure 1 shows the monthly trend in the probability of cocaine-related Google searches. Although fluctuating over time (between 869 and 1,226 per 10,000,000 searches), the general trend suggested a decline in the probability of Australian cocaine-related searches over the period. However, this trend may be affected by searches for entertainment (e.g. movies, television series or video games) about cocaine (e.g. "Cocaine Cowboys" on Netflix).

Peaks in search probabilities in Australia were observed in September 2021, March 2022 and May 2022. These were related to news media searches for Nadia Bartel (and Melbourne Storm), Shane Warne and Amber Heard, respectively. These peaks disappeared or reduced when searches on related news media were excluded. A peak was also observed in December 2021 but reflects a seasonal trend that generally occurred in most years.

By contrast, the global trend in cocaine-related searches was relatively stable over the period, and it ranged between 518 and 615 per 10,000,000 searches. The probability of cocaine-related searches was consistently higher in Australian versus global searches, suggesting cocaine-related searches were more popular in Australia.

DRUG TRENDS

Figure 1. Trend in Google searches on cocaine from July 2021 to June 2022, in Australia and worldwide.
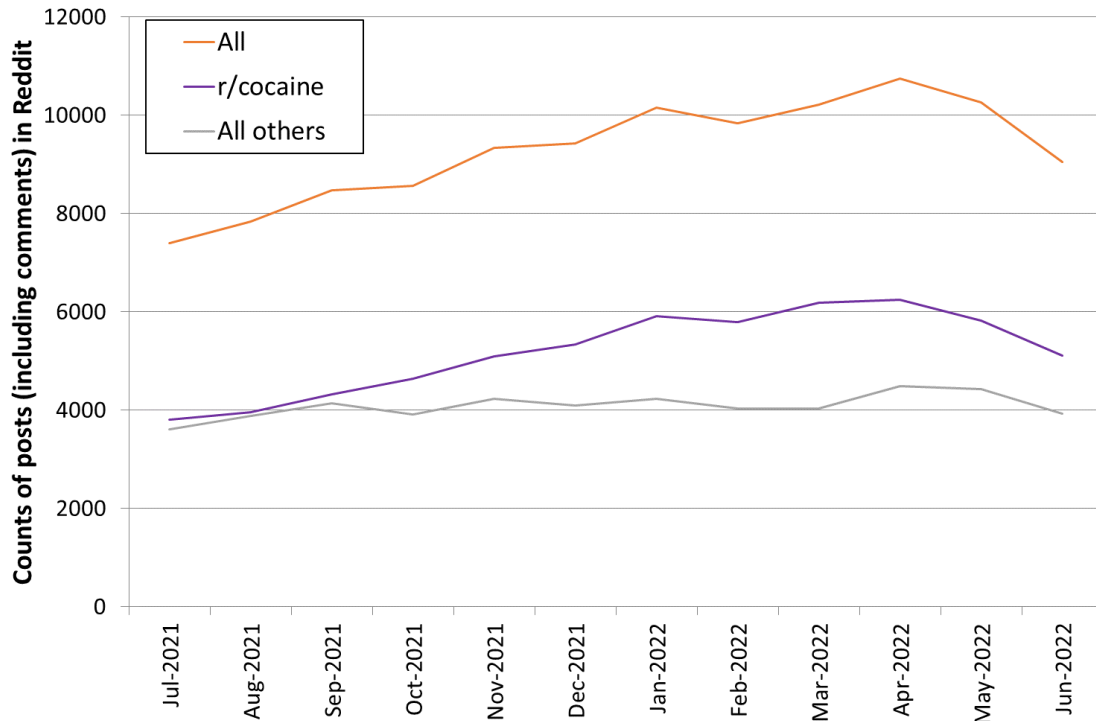

Note: *'/m/0256b'* is the freebase ID for cocaine. * search list for exclusion of popular media is described in section 3.3.1 in the Methods document.

1.2. What is the monthly trend in cocaine-related posts on Reddit?

We have identified 1,273,209 posts (including comments) about cocaine on drug-related subreddits from July 2021 to June 2022. Figure 2 shows the monthly trend in cocaine-related posts on these subreddits. Posts increased from July 2021 to April 2022, then plateaued and decreased in May and June 2022.

The increase in cocaine-related posts was largely attributable to growth in the number of posts on the r/cocaine subreddit, which also had the largest number of cocaine-related posts in the 12-month period.

Figure 2. Global trend in number of cocaine-related posts on Reddit, overall and by the major cocaine-related subreddit (r/cocaine) and other subreddits, from July 2021 to June 2022.
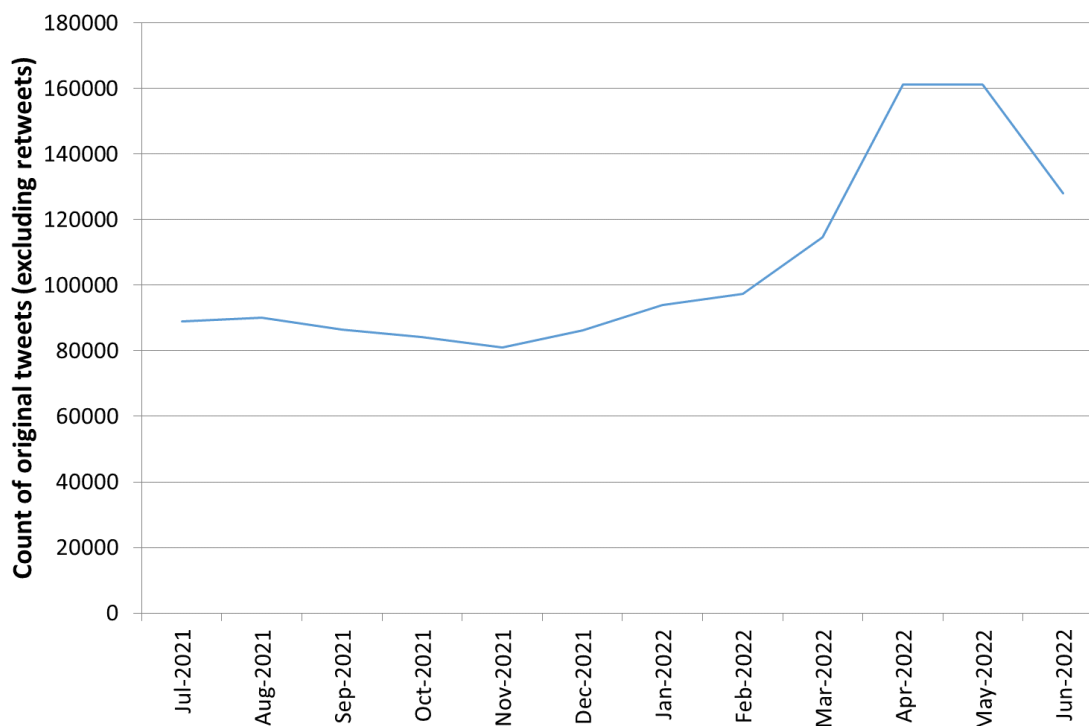

Note: The number of posts per month were rescaled to per 30-days to account for proportionally more or less posts in months with more or less days, respectively.

1.3. What is the monthly trend in cocaine-related posts on Twitter?

We have identified 3,374,330 tweets with reference to cocaine on Twitter from July 2021 to June 2022, of which 1,290,684 were original tweets (i.e., excluding retweets). Figure 3 shows the monthly trend in cocaine-related original tweets increased from around December 2021 to April 2022, then plateaued and decreased in May and June 2022.

Figure 3. Global trend in number of cocaine-related original tweets on Twitter from July 2021 to June 2022.

Note: The number of tweets per month were rescaled to per 30-days to account for proportionally more or less tweets in months with more or less days, respectively. Original tweets were derived by removing retweets from all tweets.

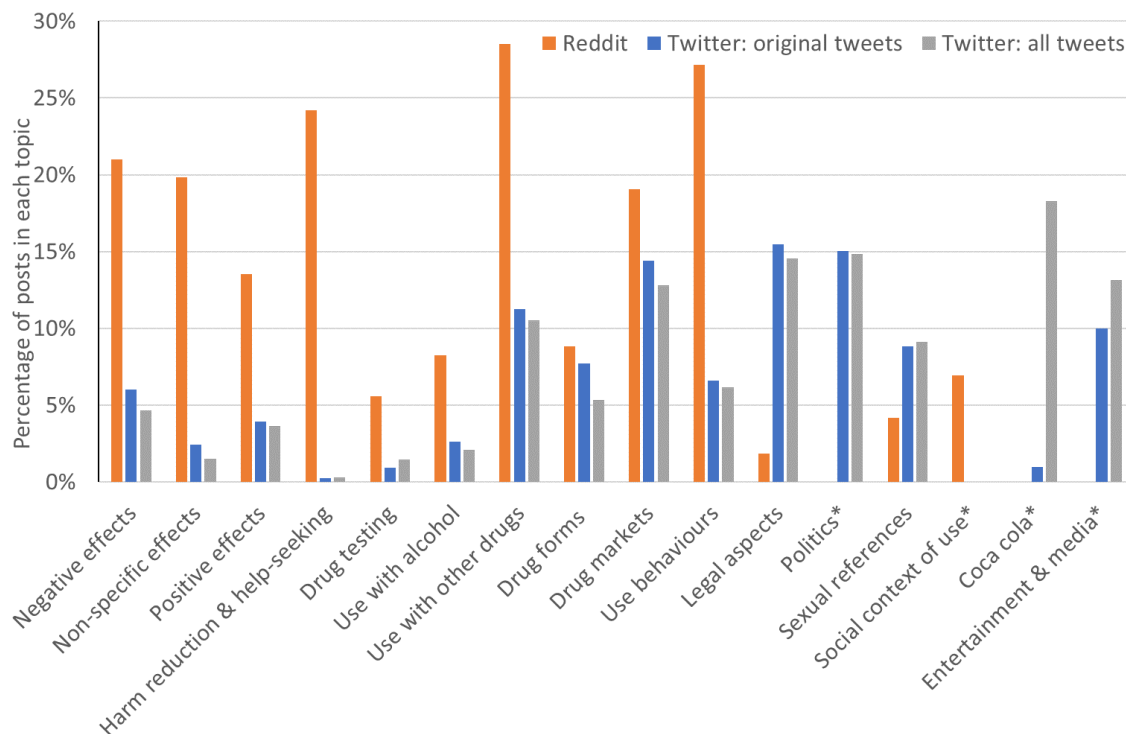## 2. Topics discussed and sentiment on cocaine-related posts in social media

2.1. What were the topics discussed in cocaine-related posts on social media?

Figure 4 shows the distribution of topics discussed in cocaine-related posts on Reddit and Twitter. Note that 80% of posts on Reddit, and 78% and 66% of original tweets and all tweets, respectively, on Twitter were categorised. A post may also be categorised to one or more topics. Please refer to Methods document for how posts were categorised to these themes. A definition of the topics and keywords used to identify them are given in Tables 2 to 4 of the Methods document.

The most common topics discussed on drug-related subreddits in relation to cocaine were use with other drugs (e.g., 'methamphetamin*'; 29%), use behaviours (e.g., 'line*', 'snort*'; 27%), harm reduction and help-seeking (e.g., 'advic*', 'water'; 24%), negative effects of use (e.g., 'anxieti*', 'die'; 21%), and drug markets (e.g., 'cut*', 'street*'; 19%). Compared with Twitter, people on drug-related subreddits were less likely to be discussing matters of law and politics, and on miscellaneous topics in relation to cocaine.

The most common topics discussed in cocaine-related original tweets on Twitter were legal aspects (e.g., 'seiz*', 'launder*'; 15%), politics (e.g., 'biden', 'republican*'; 15%) and drug markets (e.g., 'columbian*', 'bitcoin*'; 14%). Very few keywords were identified in the topics of harm reduction and help-seeking, drug testing, and use with alcohol because people were less likely to tweet about them on Twitter. While there were very few original tweets on Coca Cola (<1%), they comprised 18% of all tweets the majority of which were retweets in April 2022. For this reason, we have presented original tweets (i.e. removed retweets) for the trend in frequency of tweets in Figure 3. While there were more sexual references on Twitter as compared with Reddit, many of them were related to political tweets for the purpose of political slandering.

Figure 4. Topics discussed in cocaine-related posts on drug-related subreddits and original tweets on Twitter

Note: A definition of the topics is Table 2 of the Methods document. The numbers are the percentages of the posts categorised into a topic over all the posts. Original tweets were derived by removing retweets from all tweets. A post may also be categorised to more than one topic. *No posts were identified for the topic of politics, Coca cola and entertainment and media in drug-related subreddits because relevant terms could not be found in the top 1000 potential keyword list. Relevant terms also could not be found in the top 1000 potential keyword list for Twitter in the topic of social context of use.
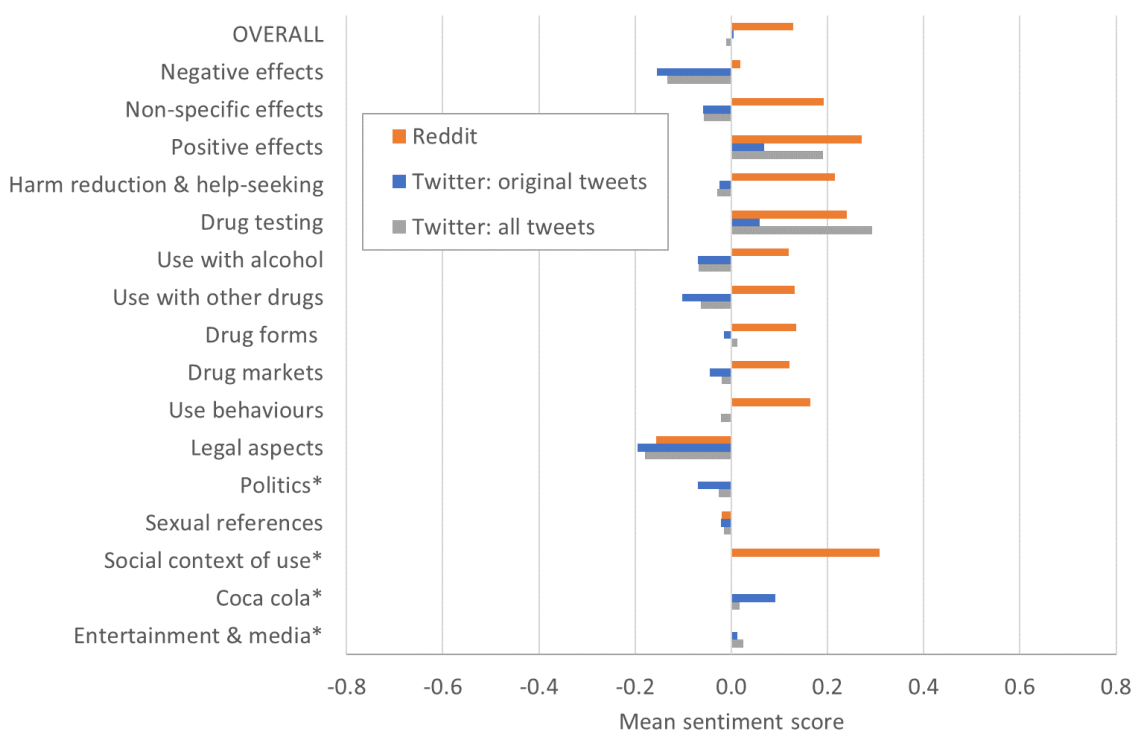
DRUG TRENDS

2.2. What is the sentiment of each topic discussed in cocaine-related posts on social media?

Figure 5 shows the mean sentiment scores of cocaine-related posts on drug-related subreddits and on Twitter. Note that a compound sentiment score is assigned to each post and that it can range from -1 to 1, with -1 being the most negative score, and 1 being the most positive score.

The discussions on drug-related subreddits are generally more positive than on Twitter. Indeed, nearly all topics received a positive compound score for Reddit except for discussions on legal aspects and sexual references. The topics which received the most positive score on Reddit were social contexts of use, positive effects of use, drug testing, harm reduction and help-seeking.

By contrast, the only topics on Twitter with a positive sentiment score were drug testing, followed by positive effects of use. Like on Reddit, tweets on legal aspects had the most negative sentiment score, and negative effects of use were also rated negatively on Twitter.

Figure 5. Mean composite score on sentiment in drug-related subreddits and on Twitter, overall and by topic

Note: Original tweets were derived by removing retweets from all tweets. *No posts were identified for the topic of politics, Coca cola and entertainment and media in drug-related subreddits because relevant terms could not be found

in the top 1000 keyword list. Relevant terms also could not be found in the top 1000 keyword list for Twitter in the topic of social contexts of use (see footnote of Figure 9).

2.3. What is the monthly trend in the broader topics discussed in cocaine-related posts on social media, and the sentiment over time?

When studying Reddit data broken down by broader topic, the percentage remains relatively stable over the period of monitoring (Figure 6). The main topic which evidenced change was the polysubstance use topic, which appeared to decline. The mean sentiment of each topic also remained relatively stable over the 12-month period (Figure 7).

Figure 6. Monthly trend in the percentage of cocaine-related posts by broader topic in drug-related subreddits.
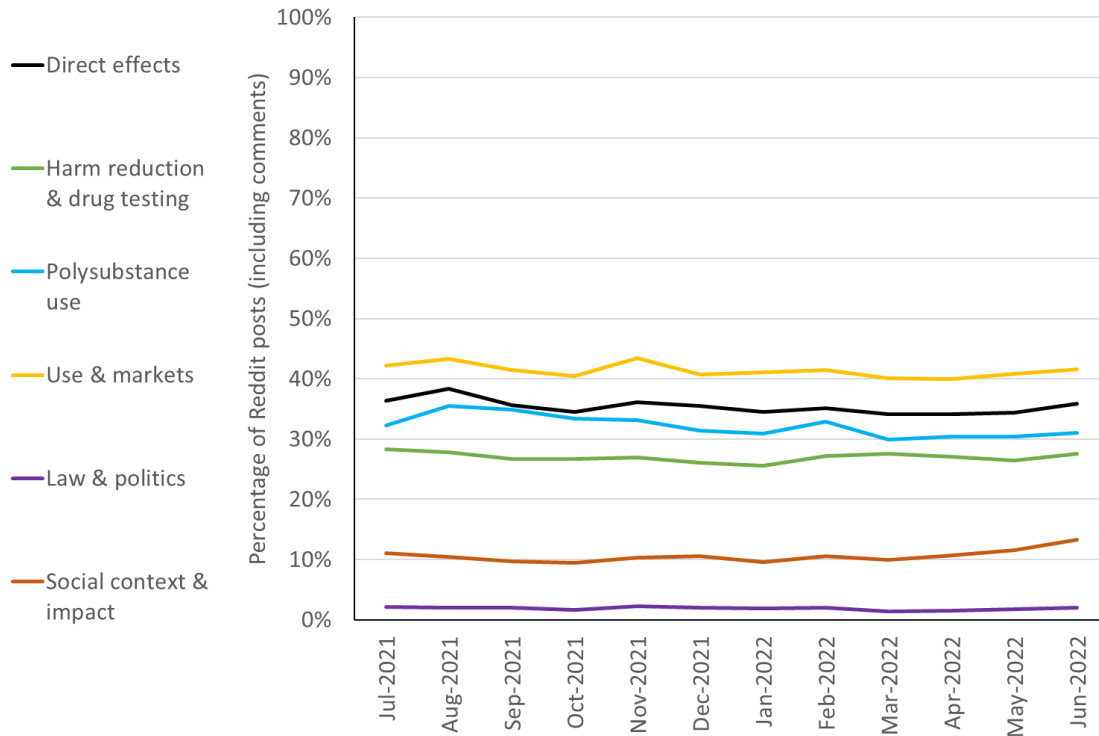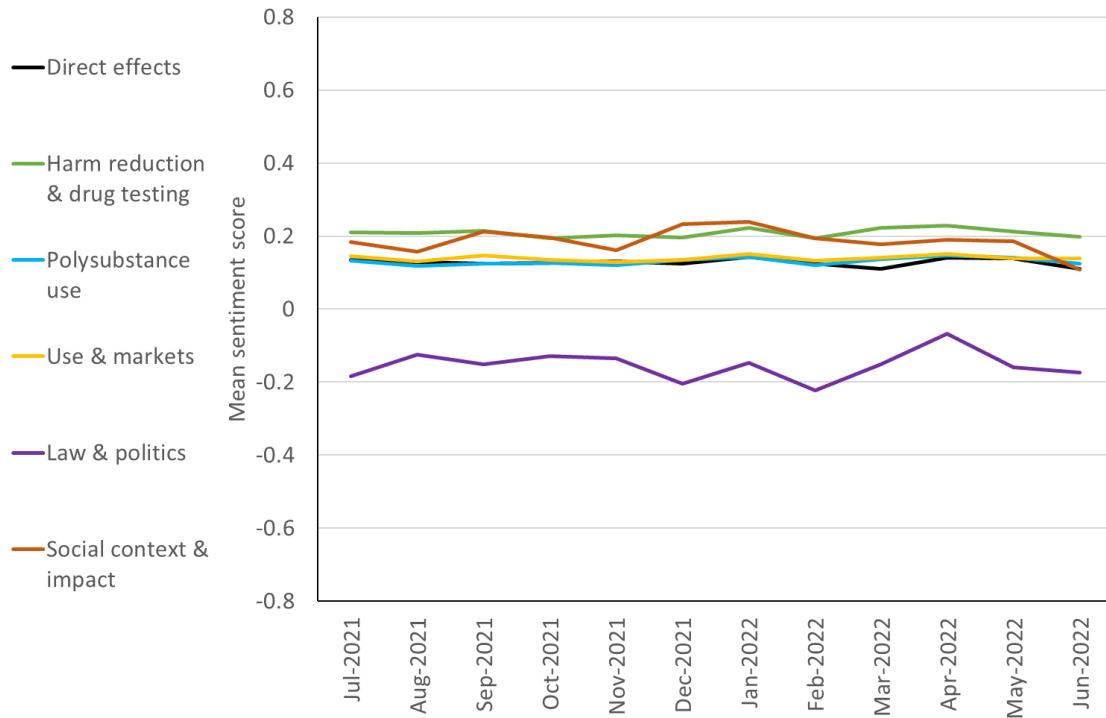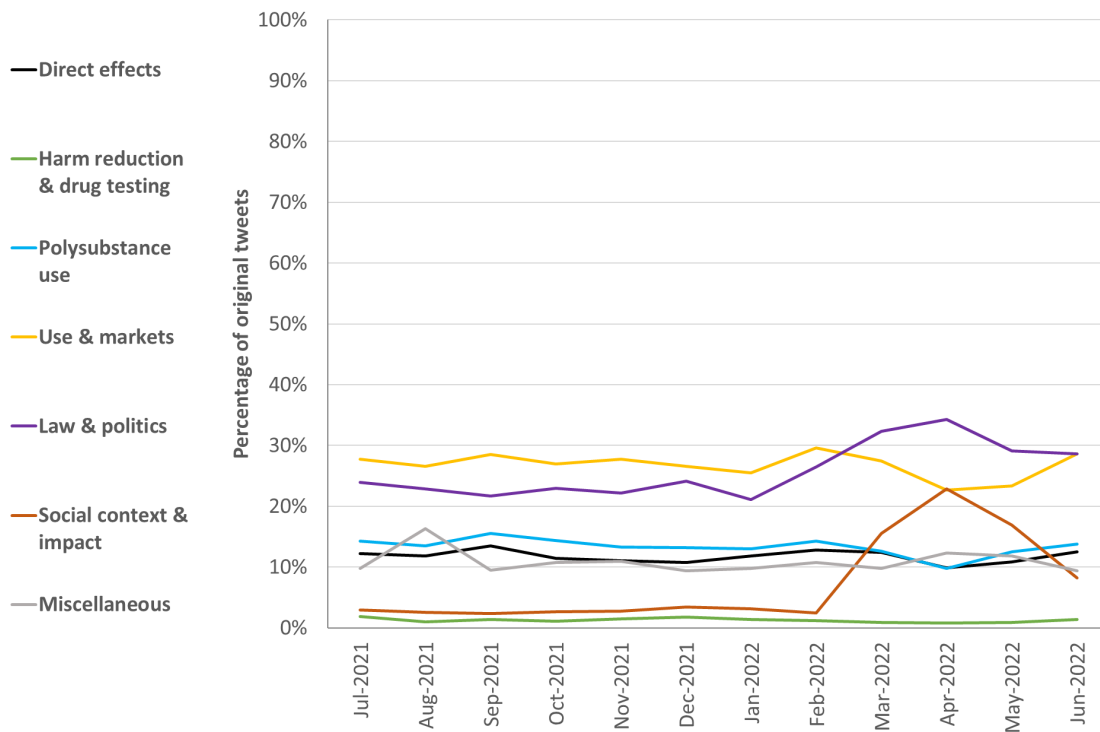
Figure 7. Monthly trend in sentiment of cocaine-related posts in each broader topic in drug-related subreddits.

With respect to Twitter, Figure 8 shows the percentage of original tweets by broader topic. The April 2022 peak in law and politics, and social context and impact of use were primarily from the increase in tweets about politics, and sexual references within these broader topics, respectively. The increase in number of tweets shown in Figure 3 is driven by the increase in the number of tweets in these topics.
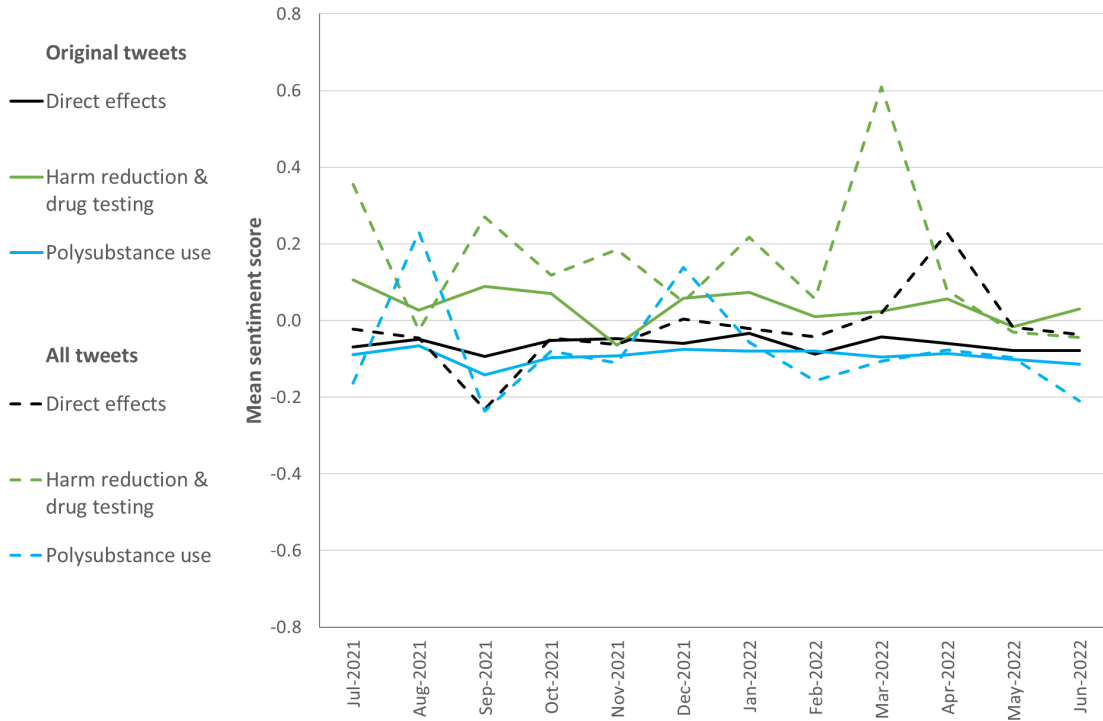
DRUG TRENDS

Figure 8. Monthly trend in the percentage of original cocaine-related tweets in each broader topic.

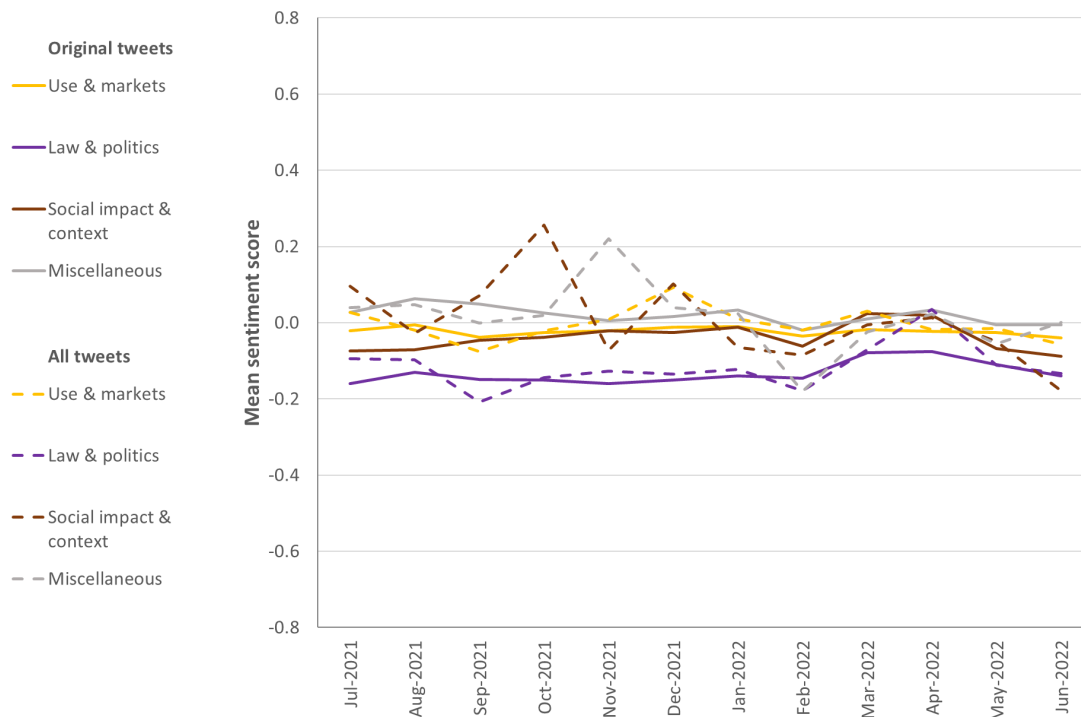Note: Original tweets were derived by removing retweets from all tweets.

Figures 9 and 10 show that the mean sentiment of original tweets in each of the broader topics appear relatively stable over time. However, these figures also show how retweets can influence the mean sentiment of particular topics over time, and that particular tweets with more divergent sentiment scores are retweeted more frequently than the original tweets. In particular, the drug testing tweets within the broader topic of harm reduction and testing had a high positive sentiment in March 2022, which was not evident when examining original tweets only. There were other less obvious spikes in positive sentiment in July and September 2021 in all drug testing tweets. There was a spike in tweets on direct effects of use, polysubstance use, and social impact and context of use in April 2022, August 2021 and October 2021, respectively, with a high positive sentiment attached to them, which were not evident with the original tweets.

Figure 9. Monthly trend in sentiment of cocaine-related tweets in the broader topics of direct effects, harm reduction and drug testing, and polysubstance use in Twitter.

Note: Original tweets were derived by removing retweets from all tweets.

DRUG TRENDS

Figure 10. Monthly trend in sentiment of all cocaine-related tweets in the broader topics of use and markets, law and politics, social impact and context, and miscellaneous in Twitter.

Note: Original tweets were derived by removing retweets from all tweets.

Future work

This is a feasibility study on the collection of drug-related data from three online sources: data on Google searches, and posts on drug-related subreddits and Twitter. It demonstrates that it is possible to collect data on an illicit drug such as cocaine by firstly collecting plausible search terms for identifying cocaine-related data, and undertaking an exploratory analysis on the subset of search terms that can be used to identify the drug.

In addition to further validation of methodology, there is scope to expand this work in multiple ways to yield important data for monitoring illicit drug trends in Australia.

Firstly, while geospatial information for Twitter data could not be obtained for this report, we will be exploring various potential approaches to identifying tweets from Australian users.

Secondly, there is also an opportunity to expand the work to other substances and other online data sources. This may be particularly useful in the monitoring of new psychoactive substances. There are limited data around their emergence, use and associated harms; social media data can likely address some of these gaps in knowledge. Similarly, there

are a burgeoning number of illicit online pharmacies on the surface web. Study of these websites could facilitate our understanding of the contribution of online surface marketplaces to diversion and counterfeit products.

Thirdly, the value of these data could be consolidated by triangulation alongside more traditional epidemiological indicators. Such work could compare findings across sources, including identifying whether social media data may act as an 'early warning' of emerging trends of concern relative to traditional epidemiological indicators which are often lagged.

Fourthly, future work could also identify the value of social media data when evaluating the impact of population-level public health interventions or other strategies to reduce drug-related harm. Ability to access historical data means that 'baseline' information can be readily accessible, and data collation does not burden participants and is comparatively less resource intensive relative to primary data collection. However, we note also that some of these platforms constantly change their terms and conditions of use including terms around the permissibility and scope of data collection allowed. In particular, there is uncertainty around how the change in ownership and management of Twitter may affect the posting of tweets and our future data collection.

Finally, more nuance in content analyses could yield important insights on topics of public health import. For example, there could be opportunities to identify strengths and gaps in knowledge of harm reduction practices to inform education endeavours. While not without ethical and methodological challenges, social media and other online data enable study of stigmatised and illicit behaviours which may not be otherwise studied in such depth, and identification of emerging trends of potential concern.

Funding

Ethical approval

This project has institutional approval and ethical approval from the University of New South Wales Human Research Ethics Committee (HC180004).

Acknowledgements

Recommended citation

Linghu, Q., Peacock, A., Bruno, R., Sutherland, R., Barratt, M. J., and Man, N. (2022). Trends in Google searches and social media discussions of cocaine, July 2021-June 2022: A pilot study. Drug Trends Bulletin Series. Sydney: National Drug and Alcohol Research Centre, UNSW Sydney. DOI: 10.26190/r4dt-aj11

Related Links:

- Cryptomarket bulletin: https://ndarc.med.unsw.edu.au/resource-analytics/trends-cryptomarket-drug-listings-oct2021-sep2022
- Cryptomarket data visualisation: https://drugtrends.shinyapps.io/cryptomarkets
- For more research from the Drug Trends program go to: https://ndarc.med.unsw.edu.au/program/drug-trends

Contact us

Email: drugtrends@unsw.edu.au