Online monitoring of New Psychoactive Substances: Reddit and online drug forums as popularity metric.

ANSHITA KHANDEWAL, 14464519, University of Amsterdam, the Netherlands
MARA SMEETS, 14629747, University of Amsterdam, the Netherlands
BORIS WOLFSON, 13551884, University of Amsterdam, the Netherlands
IMAD LOTFI, 14651610, University of Amsterdam, the Netherlands
CINTHYA CRIOLLO, 11977477, University of Amsterdam, the Netherlands

UNIVERSITY OF AMSTERDAM

ABSTRACT

The emergence of new psychoactive substances (NPS) as a replacement for classical popular drugs has led to a growing public health and law enforcement challenge in the Netherlands. Monitoring trends of NPS popularity and rapid identification of new upcoming NPS could add value to predictive policing methods. Existing methods to research NPS popularity in the Netherlands consist mainly of drug sample testing and surveying which are expensive and time-consuming. The main aim of this research is to compare the behavior of drug popularity on online forums with real-life test sample data. Correlation of online popularity indicators with real-life test sample data obtained promising results for two out of three forums tested. A predictive system for flagging new upcoming popular NPS is suggested using the best performing popularity indicator. In addition, a working dashboard prototype is created that can be used to search the online forum popularity of a NPS of interest. The dashboard shows the results of the flagging system applied on a list of current existing NPS in the Netherlands. The main limitation of this research is a lack of ground truth data. More granular data on drug popularity could be used for testing and validation of the proposed flagging system.

Keywords New Psychoactive Substances, NPS, drug popularity, drug forum, Reddit, online monitoring

1 INTRODUCTION

In 2008-2009, a 'dip' in the ecstasy market due to a shortage on MDMA caused New Psychoactive Substances (NPS) to be used as a replacement for classical popular drugs [7]. However, the NPS soon started to be purchased as specific drugs of use, and subsequently a new NPS market arose. NPS are sometimes also referred to as 'legal highs', 'designer drugs' or 'research chemicals'. Most NPS have a similar effect as traditional illegal drugs, but are not (yet) covered by drug legislation. The rate at which new NPS emerge causes difficulties for existing drug control systems. In addition, the online marketing and sales of NPS has increased substantially. The ease with which NPS are distributed and purchased online has resulted in a growing public health and law enforcement challenge [22]. Minor modifications of the substance structure are easily performed ensuring that existing legislation is bypassed [12].

As a response to this rapidly growing new NPS market, the Dutch Trimbos institute set up a hotline for new drugs (Meldpunt Nieuwe Drugs). This institution specifically gathers information on NPS, as part of the existing Drugs Information and Monitoring System (DIMS) and the National Drug Monitor (NDM) [17]. In addition, the European Monitoring Centre for Drugs and Drug Addiction (EM-CDDA) is the leading authority on drug monitoring in the European Union. Participating countries such as the Netherlands notify the EMCDDA when a new NPS is discovered. At the end of 2021, the EMCDDA was monitoring around 880 NPS, 52 of which were first reported in Europe in 2021 [13].

In 2019, a European directive was implemented in the Dutch Opium Act in order to include NPS under the definition of 'drug', making it possible for them to be declared illegal faster [27]. The Dutch government is working on a total ban on all NPS [11]. Still plenty of uncertainties exist about the effects of this policy on reducing public health risk and criminality. In 2016 the British government adopted a total ban. It does appear that vendor shops have disappeared, but it is unclear whether this is a gain or a loss [9]. Ireland has had a total ban in place for five years and although vendor shops disappeared and less individuals were being treated for NPS dependence or abuse, a survey found that 22% of 15–24-year-olds reported using NPS, compared to a European average of 8% [10]. Therefore, rapid identification of new upcoming popular NPS is still needed in order to prevent public health risks and to adequately respond in terms of legislation.

Rapid identification of new upcoming popular NPS is a form of predictive policing. In predictive policing, data and analytical techniques are used to identify areas and individuals at high risk of criminal activity to deploy law enforcement more effectively. Existing methods of the Trimbos Institute to research drug popularity consist mainly of drug sample testing and surveying (online or at festival/party locations) [11]. The main indicator for the popularity is the annual count of handed-in drugs at the test centres. The Netherlands is one of the few countries that offers the opportunity to anonymously hand in drug samples to test substances free-of-charge. The popularity indicator can be retrieved from what people think to have bought, regardless of the actual substance [11].

It should be noted that the data from the Trimbos institute on NPS popularity is extremely scarce and is only reported in annual reports.

Furthermore, large surveying is time-consuming and expensive. Given the cost and time taken in real-life drug testing, the value of online sources in early warning systems for popular NPS should be examined. Monitoring online forums could be used as a close to real-time monitoring system in contrast to the annual report of the Trimbos Institute. In addition, it could be significantly less costly and time-consuming. In recent years, researchers and law enforcement agencies have begun to explore the use of data scraped from social media platforms as a way to improve predictive policing methods [2]. The aim of our research is to test popularity indicators of Reddit and Dutch drug forums against the popularity indicators of the Trimbos Institute. Moreover, this report aims to create a predictive system of the best performing source in order to improve the existing early warning system of popular NPS. Therefore, the following research questions are answered:

(1) *Do online forum sources mirror the behaviour of real-life test sample data of NPS popularity in the Netherlands?*
(2) *To what extent can online forum sources be used for predicting new upcoming popular NPS in the Netherlands?*

2 LITERATURE REVIEW

The evolution of predictive policing has been marked by a shift from traditional methods to more advanced and data-driven techniques. Traditional methods, while useful, had several limitations such as being based on past crime data and it required a significant amount of manual data analysis, which was time-consuming, expensive and resource-intensive [1]. The use of data from online sources such as social media platforms allows for more real-time monitoring of criminal activity, as well as the ability to analyze a wider range of factors that may be related to crime [2]. This data can also be used to identify emerging trends and patterns that may not be apparent from traditional crime data alone, as shown by the effectiveness of predictive policing in previous studies [3–6].

A recent method of tracking the emergence of NPS is through the use of data scraping techniques applied to online drug forums. One study researched the number of replies to threads on drugs-forum.com as popularity indicator and compared the first occurrence dates on the forum to the first notification dates in the EMCDDA [24]. A more recent study analyzed the text of more than 67 sub-Reddits and assessed the relationship between online mentions of drugs and their involvement in toxic exposures or overdoses in the USA [23]. Overall, these studies suggest that data scraping techniques applied to online drug forums can be a useful tool for predicting new trends in NPS. Another study also concluded that monitoring forums is useful to complement existing NPS monitors that focus on developments in the drug market and have possibly a predictive value with respect to new trends in the NPS-market that are not yet identified [8]. However, it is important to note that further research is needed to establish the full potential of this method. The connection between conversations on drug forums and drug-user behaviors is not widely researched, however it is reasonable to argue that monitoring such discussions could help policy and law enforcement agencies to identify emerging trends in NPS use and markets [19].

Understanding consumers' purchasing behavior and perceptions is key to successfully designing and implementing public health interventions or law enforcement strategies [21]. In 2019, large-scale European research showed that 16.9% of the NPS users bought their drugs from online webshops [12]. NPS are sold across the web, and information about consumption is shared by online communities through drug forums, YouTube channels, social networks, and smartphone applications [18]. People are no longer restricted to local drug dealers as (il)legal drugs can simply be ordered online. The dark net facilitates anonymous buying and selling of NPS, including to very young people, in ways that avoid law enforcement [18]. However, the dark web is more commonly used for purchasing traditional drugs, while the clear-web is preferred for NPS [12; 21].

On these clear-webshops, NPS are often sold online as 'legal' and 'safer' than traditional drugs with smart marketing strategies [20]. NPS are sold openly as well as hidden by several aliases, such as bath salts, comics, movies, landscapes/space or the expected psychotropic effects. Other advertising methods, often implemented to especially attract young people, included NPS with colorful appearance and popular imprints.

Previous research also shows that there is an increase in the drug trade in social media platforms, such as Telegram, which accelerated at the beginning of the COVID-19 pandemic, and kept present afterwards [29]. Telegram offers anonymous login, sign-up without sim card and auto-delete messages for all participants without a trace, thus it gives users the possibility of fully owning their entire message history and having ultimate privacy [31]. Hence, Telegram is popular among criminals to trade drugs with a high level of anonymity [30].
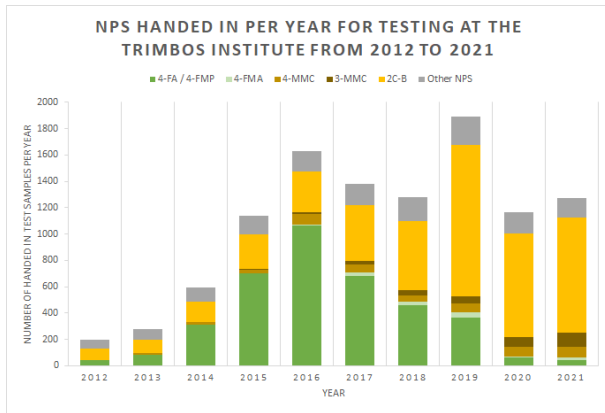
3 METHODOLOGY

A link to the GitHub containing all relevant code and sources can be found in the header of the report.
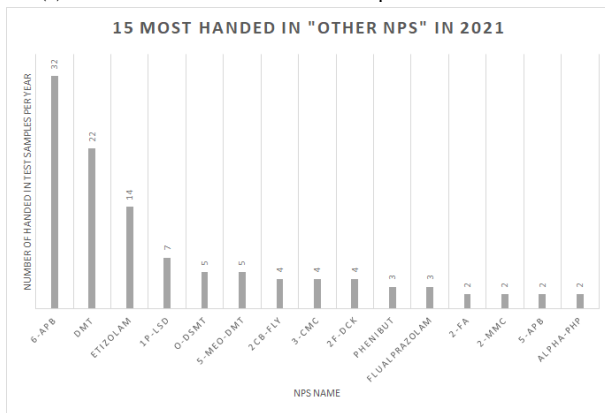
3.1 Data collection

The Trimbos institute is the main authority in the Netherlands that collects and analyzes data on the drug market. Since 1993, the DIMS institute offers anonymous and free drug sample testing and the number of handed-in test samples has been steadily increasing [17]. Apart from more traditional drugs, 2C-B, 3-MMC, 4-MMC (Mephedrone), 4-FA (4-FMP), and 4-FMA have been separately mentioned in their annual reports since 2012 (Figure 1a). Unfortunately, the Trimbos institute has collected data on 'other NPS' only since 2019. For 2021, the top 15 most handed-in 'other NPS' are shown (Figure 1b).

3.1.1 Vendor webshops. As a significant portion of NPS is bought via clear webshops, the online vendor supply of NPS gives a good overview of the current NPS market in the Netherlands (Table 1). In addition, it is the sole openly available source. The first five relevant vendor webshops on the Google search engine are scraped using Python's BeautifulSoup package to collect data on their NPS supply. The NPS list resulting from the web scraper is used for prediction of upcoming new popular NPS. The following webshops are scraped:

- funcaps.nl
- research-chemicals-kopen.com

(a) Number of handed in NPS test samples from 2012 to 2021.

(b) 15 most handed in 'other NPS' test samples in 2021.

Fig. 1. Trimbos institute NPS test sample data [11].

- realchems.nl
- professor.nl
- 4fmp.net

Source	Percentage Share
Friends	51.3
Webshops	16.9
Private Dealers	12.4
Street Dealers	9.1
Regular Shops	8.4

Table 1. % share of drugs sold by different sources in Europe, as researched in 2019 [12].

3.1.2 Reddit. Reddit is chosen due to it's tremendous user base which correpsonds to approximately 1.8B monthly visits. All relevant sub-Reddits (Appendix 7.1) are scraped based on past research on the topic [23]. Relevant popularity indicators from Reddit are the number of mentions (of the keyword), the number of comments, and the number of upvotes (as seen in Table 2). As keywords are

often also mentioned in URLs contained in Reddit posts, these are removed from the text body.

Forum	Popularity Indicator
Reddit	Mentions
	Comments
	Upvotes
Drugsforum.nl	Views
	Comments
	Mentions
Drugsforum.info	Views
	Comments
	Mentions

Table 2. Popularity indicators for each forum.

3.1.3 Dutch drug forums. Two main Dutch drugforums, drugforums.nl and drugforums.info, with approximately 350K and 58K monthly visits respectively, are scraped using Python library BeautifulSoup. The categories scraped for drugforums.nl are 'Research Cshemicals', 'Drugs General' and 'Trip Reports', while only 'Research Chemicals' was scraped for drugforums.info. Relevant popularity indicators stored in the database for these forums are the number of mentions, number of comments, and number of views (Table 2). Unlike Reddit, upvotes are not available.

3.1.4 Other datasets explored. Other open-source datasets on drugs are available, such as the Dutch court cases database, or news articles. In the initiation phase of this research, data from these sources was collected. However, as NPS are mostly not included in drug legislation, and as many of them occur in the market for short periods, NPS are unlikely to be frequently mentioned in court cases and news articles to give relevant data on NPS popularity.

Additionally, Telegram chats are explored for this report. The chat history from two selected Telegram chats were downloaded as json files. The first chat offers drugs for sale among the trading of other products, such as cryptocurrency investments. The purchase is done in a private chat, redirected from the main group channel. The size of the group exceeded hundreds of MB, which is not possible to analyze on an average laptop processor. Dividing the data into smaller datasets and zip compression methods were tried. Nevertheless, this did not offer optimal results, thus the chat was dropped from the research.

In the other chat, opinions on drug micro dosing are discussed. The number of mentions of relevant NPS were analyzed. Nonetheless, the micro dosing chat did not show relevant results. To sum up, the investigation of Telegram did not hold significant findings. However, further investigation is recommended as it could prevent Telegram from being a major platform in the drugs trading market in the future [28].

3.2 Statistical analysis

3.2.1 Search function. In order to find all the relevant mentions for the search drug in the corpus, a regular expression (i.e. regex) approach was implemented. The expressions were designed in such a way that all the combinations of hyphens and spaces between the

digits and letters would be included. For example the search for the keyword '123-abc-456', will also look for 123abc456, 123 abc 456, 123-abc 456, etc. In addition, a dictionary of known drug synonyms was designed, if a drug has known synonyms, the search function, in addition to the keyword, will return also mentions of its synonyms (Table 3). Only known synonyms for the 20 most handed in NPS are included.

Drug	Synonyms
4-FA	4-FMP
4-MMC	Mephedrone, Mefedron
6-ABP	Benzofury, Benzo Fury
Phenibut	Fenibut

Table 3. Drug synonyms for search.

3.2.2 Popularity indicators. The performance of the popularity indicators of Reddit, drugsforum.nl and drugsforum.info are tested for the five main NPS for 2012 to 2021. It should be noted that the Trimbos test sample data only includes yearly data, whereas Reddit, drugsforum.nl and drugsforum.info contain monthly data from the date of the first mention to January 2023. In order to analyze the trend over a larger period of time, and in order to deal with monthly spikes in online activity, the monthly data is smoothed using an averaging filter. Two methods are tested for smoothing: 1) The monthly data of the smoothed plot is summed per year. 2) Piecewise linear regression with 12 breaking points is performed on the smoothed time series. The actual summed monthly data, as well as both methods for smoothing are compared with the Trimbos test sample data from 2012 to 2021 using Pearson correlation.

3.2.3 First notification dates. The first notification date from the EMCDDA is retrieved for the 20 most handed in NPS in 2021 (Table 6). The value of online forums for rapid notification is assessed by comparing the first occurrence dates of the NPS on the forums to the first notification dates of the EMCDDA.

3.3 Flagging system

The best performing popularity indicator and online forum source is used for the predictive system. Flagging is performed based on three indicators: 1) Most increase in popularity indicator in the past six months. 2) Highest value of the popularity indicator in the past three months. 3) Newly occurring in the past three months. Prediction of upcoming popular NPS is applied on the list of NPS that is scraped from the vendor webshops (Appendix 7.3).

3.4 Final product

The final product is a dashboard that can be used by the police academy. Entering a NPS search word gives the best performing popularity indicator time series for the online forum sources. In addition, based on the scraped NPS list from vendor webshops, the top 5 NPS that resulted from the flagging system are presented.

3.4.1 User validation. In order to analyze the usability of the dashboard, summative research is conducted using a survey. The survey is performed independently without any help of the supervisor.

During the evaluation session, the users were asked to interact with the dashboard interface. Afterwards, their user experience was requested using the survey (Appendix 7.2). In total, the survey consists of seven questions representing common tasks that could be performed in the dashboard. When participants ask questions or hesitate, the task is marked for further examination. At the end of the survey, participants are asked to give additional suggestions about the tool.

4 RESULTS

4.1 Data collection

The vendor webshop scraper on January 18th (2023) resulted in a list of 187 currently available NPS (Appendix 7.3). All five main NPS are identified on the vendor webshops. From the 15 mostly handed in 'other NPS' in 2021, only DMT is not identified on the vendor webshops. One of five webshops, funcaps.nl, does sell DMT however. It does not sell it specifically as DMT, but as active component in 'Jurema Indian Spirit'.

4.2 Statistical analysis

4.2.1 Popularity indicators. The results of the Pearson's correlation and their corresponding p-value between the popularity indicators from Reddit, drugsforum.nl and drugforum.info and the Trimbos institute's test sample data are presented in Table 7. As explained in Section 3.2.2, the popularity indicators are tested for the five NPS: 2C-B, 3-MMC, 4-MMC (Mephedrone), 4-FA (4-FMP) and 4-FMA. Overall, the moving average of the number of views shows the highest average correlation for drugsforum.nl (Pearson's R 0.70). In particular, 3-MMC performs well as all correlation values of popularity indicators from drugsforum.nl range between 0.88 and 0.94. The plots for the number of views of 3-MMC on drugsforum.nl show significant similarity to the Trimbos test sample data (Figure 2a). The moving average of the number of views of 3-MMC corresponds to a Pearson's R of 0.92 with p-value 1.8e-4 (Table 4). Moreover, the number of comments and mentions for 2C-B perform considerably well, as do the number of views and mentions for 4-FA.

For Reddit, the number of mentions popularity indicator shows the highest average correlation values (Pearson's R 0.53). Specifically 2C-B and 3-MMC show high correlation values on all indicators, whereas 4-FA and 4-FMA have extremely low correlation values. The highest correlation values again refer to 3-MMC, ranging from 0.84 to 0.89, where the actual data of the number of mentions corresponds to a Pearson's R of 0.89 with p-value 5.5e-4 (Table 4). This is also reflected in the plot (Figure 2b). It should be noted that 4-FA performs significantly worse on Reddit than in drugsforum.nl.

Moreover, the results show that the correlations for drugsforum.info are extremely low and mainly negative. The highest correlation is seen for the piecewise linear regression of the number of mentions (Pearson's R 0.25, p-value 0.48). However, the results show high (negative) correlation values for 2C-B and 4-MMC, namely between -0.69 and -0.87.
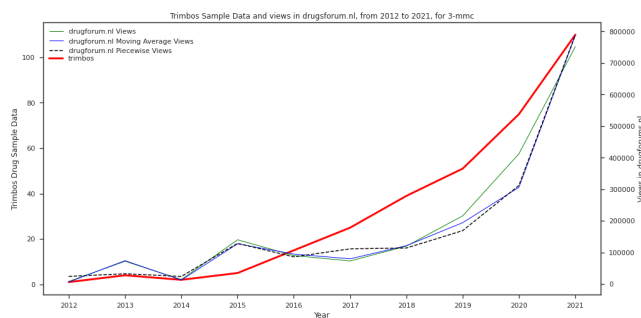
4.2.2 Flagging system. The flagging system for predicting new popular NPS is applied on the NPS list that is scraped from the vendor webshops. The results of the Pearson's correlation analysis show that on average the number of views on drugsforum.nl appears to

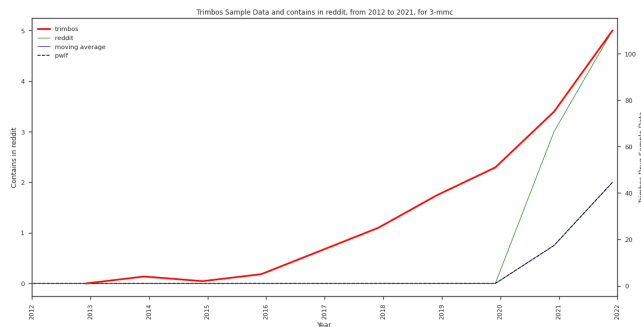| NPS	drugsforum.nl		Reddit	
	Correlation	p-value	Correlation	p-value
2C-B	0.58	0.082	0.79	0.006
3-MMC	0.92	1.8e-4	0.89	5.5e-4
4-MMC	0.69	0.026	0.69	0.026
4-FA	0.82	0.004	0.15	0.693
4-FMA	0.49	0.150	0.14	0.679

Table 4. Correlation & p-values of popularity metrics with highest average correlation with Trimbos test sample data, i.e. rolling average views for drugsforum.nl and yearly mentions for reddit.com (as seen in Table 7).

Top 5	Largest increase in the past 6 months	Largest number of views in the past 3 months	Newly occurring in the past 3 months
1	2-MMC	3-MMC	4-ACO-DET
2	4-FMPH	MET	A-D2PV
3	2-FMA	3-CMC	
4	2F-DCK	6-APB	
5	2C-C	LSD	

Table 5. Results of the flagging system applied on the current list of NPS.

4.2.3 First notification dates. Table 6 shows the first notification dates of the EMCDDA early warning system of the 20 most popular NPS in 2021 according to the Trimbos institute. In addition, the month and year of the first occurrence of the NPS in the forum sources are shown. The values that occur in the forum source prior to the first notification of the EMCDDA are indicated in red. The results show that 8 out of 20 NPS occur on forums prior to the first notification date of the EMCDDA. For 4 out of 20 NPS, no date was found in the EMCDDA reports.

4.3 Final product user validation

4.3.1 Choice-based Questions. The survey resulted in eight respondents. On average, the participants were able to finish all tasks within seven minutes and, on average, each task was correctly completed by 80% of the participants. Users emphasized that the dashboard was easy to use and informative. They also indicated that it was easy to interact with the search bars and year filter tool. No broken links or errors in the interface were discovered. Three participants indicated that the dashboard was slow in showing results. Loading time, however, can be attributed to the limited capacity of the laptop processor used to perform the surveys. It is assumed that the end users have sufficiently powerful processors and higher-speed internet to run the dashboard.

4.3.2 User Suggestions. Two users suggested to change the title of the 'Top 5 newly occurring NPS' to 'Top newly occurring NPS' since only two drugs are currently displayed in the list. The tool is programmed to display the top five results, regardless of how many NPS are newly occurring. Nonetheless, the title was changed to make the dashboard more understandable. Additionally, the participants suggested to add a loading icon when results are being displayed. This suggestion is not implemented due to time constraints, however, further development of the dashboard could include the a loading icon.

5 DISCUSSION

5.1 Limitations

5.1.1 Ground truth data. The Trimbos institute data offers some limitations, the primary of which is data availability. In order to research the popularity of NPS, Trimbos uses the records of their test sample data, which is reported for yearly time periods. The popularity indicators from the drug forums offer more granular data (i.e. monthly data), thus smaller trends are not accounted for in correlation, and it is not possible to how the popularity indicators

(a) Number of views of 3-MMC on drugsforum.nl.

(b) Number of mentions of 3-MMC on Reddit.

Fig. 2. Yearly summed plots for 3-MMC of the number of views on drugsforum.nl and the number of mentions on Reddit.

be the best performing source. Therefore, this source is used for the flagging system. The flagging system leads to a top 5 NPS per category (Table 5). It should be noted that our flagging system for rising nps has not been verified, as discussed in Section 5.1.3.

NPS name	EMCDDA	Reddit	drugsforum.nl	drugsforum.info
2CB-FLY	15-2-2007	Nov-21	Oct-09	Not found
4-MMC (mephedrone)	7-3-2008	Mar-16	Dec-08	Jan-11
4-FA (4-FMP)	1-12-2008	Feb-17	Aug-07	Jan-11
4-FMA	24-3-2010	Jan-21	May-12	Jan-13
5-APB	4-12-2010	Aug-17	Jan-11	Jan-11
Alpha-PHP	4-4-2011	Oct-20	May-15	Jun-14
6-APB (benzofury)	30-6-2011	Jun-16	May-10	Jan-11
Etizolam	2-12-2011	Jan-14	Apr-12	Jun-11
2-FA	21-2-2012	Nov-17	Jan-09	Apr-12
Phenibut	8-3-2012	Jan-16	Mar-10	Apr-11
3-MMC	5-9-2012	Nov-18	Jan-12	May-12
2-MMC	8-5-2014	Nov-22	Jan-20	Jul-15
3-CMC	14-10-2014	Nov-20	May-15	Sep-16
1P-LSD	6-3-2015	Jan-16	Mar-15	Jan-15
2F-DCK	6-10-2016	apr-17	feb-19	Not found
Flualprazolam	22-11-2019	May-20	Apr-19	Jul-18
2C-B	Not found	Jun-09	Mar-05	Mar-11
DMT	Not found	Dec-11	Sep-03	Mar-11
O-DSMT	Not found	May-19	Jan-19	Not found
5-MeO-DMT	Not found	Dec-13	Jan-06	Jun-14

Table 6. First notification dates by EMCDDA of the top 20 most handed in NPS at Trimbos institute test centers in 2021, and first occurence dates in Reddit, drugsforum.nl and drugsforum.info [14–16]. All dates that indicate earlier notification than the EMCDDA date are shown in red.

mimic real-life data within a year. In addition, the Trimbos institute has only started to report the number of test samples that are classified as 'other NPS' since 2019. From this class, merely the top 10 are reported. This means that a significant amount of potentially useful data is lost in their records.

Moreover, the data of the test samples applies only to NPS consumers that decide to hand in drugs for testing. Bias exists here, as this refers to a specific type of consumer only. It should be noted that the usage, and thus popularity, of all other consumers who do not hand in test samples cannot be determined from this indicator. For this reason, the Trimbos institute has performed large surveys on drug use in the Netherlands (Appendix 7.6). Unfortunately, the survey was only performed in 2013, 2016 and 2020, which is again inadequate data availability for statistical analysis. In addition, this bias applies to different kinds of drugs as well. People are more likely to test their drugs if it is new or if they are not confident in the quality or seller, leading to long-used drugs being underrepresented in the test sample data. Hence, the Trimbos data fails to represent the popularity of these drugs, in a way that our metrics do not.

5.1.2 Online forum sources. Unlike the Dutch drugs forums, Reddit's user base is not limited to Dutch users only. Therefore, using Reddit as a popularity metric for Dutch NPS popularity could be dubious. However, international forums could potentially also influence Dutch popularity. Moreover, the plots of the number of mentions on Reddit seem to incorporate a time shift with respect to the Trimbos test sample data (Figure 4). In particular, the plots for 4-FA and 4-FMA suggest a time shift of approximately one and two years respectively. This could be the result of NPS first occurring in

the Netherlands before being moved (i.e. trafficked) to other countries. Further analysis using cross-correlation could be performed in order to research this hypothesis.

Furthermore, popularity indicators such as the number of views are inherently biased when the total number of users of a forum increases (or decreases) over time. For example, as explained in Section 3.1.3, drugsforum.nl has six times the amount of monthly visits compared to drugsforum.info. If forum popularity changes and all users of drugsforum.nl were to move to drugsforum.info, this would affect the results significantly. A possible improvement of this research would be to actively monitor forum popularity as well, or normalize over the number of users at any given time. However, due to time constraints this is not incorporated in this research. Another possible improvement would be the active search for, and monitoring of, new emerging platforms where drugs are discussed. The methods of this research could be additionally be tested and validated on these platforms.

5.1.3 Flagging system. The main limitation of the flagging system is that currently no validation data exists. The flagging of new upcoming popular NPS is merely based on the authors' intuition on the popularity indicators. More extensive data collection on monthly popularity needs to be performed in order to test and validate the flagging system methods. More regular surveys similar to the ones performed by the Trimbos institute (Appendix 7.6) are a good example. This would lead to more granular data to validate the predictions from the proposed flagging system.

Moreover, the flagging system is applied on a list of NPS that is entirely based on scraping of vendor webshops. As explained in Section 1, a total ban on NPS could cause the distribution of NPS to move to the dark web, or elsewhere. Consequentially, there would

be no, or little, real-time data on the illegal NPS market, causing a flaw in the proposed flagging system. We believe, however, that our system could be meaningful to the police until a ban would be in place. Especially if other methods would be incorporated that inventorize the current NPS market.

5.2 Ethical responsibility

It should be noted that the data sources used in this report consist of privacy sensitive data. Reddit, drugsforum.nl and Telegram collect (some) personal data. However, the terms and conditions for these open sources state that the only publicly available sensitive data are the user names. User names can, however, be traced back to privacy sensitive end-products. As user data is not needed for our research, all user names are disregarded. It is important to realize that this research is meant for law enforcement authorities only, and that the data should be used with the highest level of responsibility.

6 CONCLUSION

To summarize, this research performed Pearson's correlation between popularity indicators on online drug forums and real-life test sample data from the Trimbos institute. The results showed that the number of views on drugsforum.nl has the highest average correlation with the Trimbos data for 2C-B, 3-MMC, 4-MMC, 4-FA and 4-FMA. Using this popularity indicator to monitor the popularity of NPS and for predicting new upcoming popular NPS could add significant value to predictive policing methods, prevent public health risk and contribute to law enforcement. Using the best performing metrics from this research, a flagging system was created to mark three types of NPS: rising in popularity, newly occurring, and currently most popular. A dashboard application was created to show the relevant results to the police.

6.1 Future recommendations

A lack of ground truth data on NPS popularity limits validation of the results. More regular surveying and more concise data collection of the Trimbos institute could lead to improvement of our system. In addition, our research incorporates regular expressions and some common synonyms of NPS. The system could however be improved by adding more synonyms, for example by scraping synonyms from psychonautwiki.com. This, as well as including data updating opportunities, could improve the usability of the dashboard significantly.

Moreover, the value of Reddit as relevant international source could be further examined using methods like time-shifted cross-validation. The results for Reddit seem to indicate a time shift compared to the Dutch drugsforum.nl. Further research on this topic is recommended. In addition, research on how to effectively use Telegram chat data could be of great value. Exploration of Telegram chats revealed the presence of a trading market of NPS, thus future investigation is recommended as the platform may develop into a major source to sell drugs in the future.

REFERENCES

[1] Mugari, Ishmael & Obioha, Emeka. (2021). Predictive Policing and Crime Control in The United States of America and Europe: Trends in a Decade of Research and the Future of Predictive Policing. Social Sciences. 10. 234. 10.3390/socsci10060234.

[2] Brett Drury, Samuel Morais Drury, Md Arafatur Rahman, Ihsan Ullah, A social network of crime: A review of the use of social networks for crime and the detection of crime, Online Social Networks and Media, Volume 30, 2022, 100211, ISSN 2468-6964, https://doi.org/10.1016/j.osnem.2022.100211

[3] Perry, Walter L.; Brian McInnis, Carter C. Price, Susan C. Smith, and John S. Hollywood. 2013. Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations. Washington, DC: RAND Corporation

[4] Bachner, Jennifer. 2013. Predictive Policing: Preventing Crime with Data and Analytics. Washington, DC: IBM, Centre for the Business of Government, pp. 86–90.

[5] Turner, George, Jeff Brantingham, and George Mohler. 2014. Technology Talk: Predictive policing in action in Atlanta, Georgia. The Police Chief 81: 72–74.

[6] Ferguson, Andrew. 2020. Predictive Policing Theory. Washington College of Law Research 24: 2020–10.
Richardson, Rashida, Jason M. Schultz, and Kate Crawford. 2019. Dirty Predictions: How civil Rights violations impact police data, predictive policing and justice. New York University Law Review 94: 193–64

[7] Hondebrink, L., Nugteren-van Lonkhuyzen, J. J., Van Der Gouwe, D., & Brunt, T. M. (2015). Monitoring new psychoactive substances (NPS) in The Netherlands: data from the drug market and the Poisons Information Centre. Drug and alcohol dependence, 147, 109-115.

[8] Daan Van der Gouwe, Matthijs Blankers, and Margriet Van Laar. Online drug monitoring - trimbos.nl. Retrieved January 13, 2023 from https://www.trimbos.nl/docs/27577052-08c4-42c3-bc93-e510e4ee2258.pdf

[9] Reuter, P., & Pardo, B. (2017). New psychoactive substances: Are there any good options for regulating new psychoactive substances?. International Journal of Drug Policy, 40, 117-122.

[10] Reuter, P., Pardo, B. (2017). Can new psychoactive substances be regulated effectively? An assessment of the British Psychoactive Substances Bill. Addiction, 112(1), 25-31.

[11] Nationale Drug Monitor, editie 2023. NPS 8.8.2 Samenstelling NPS monsters - Nationale Drug Monitor. *https://www.nationaledrugmonitor.nl/nps-kwaliteit/*. Accessed on: 10 januari 2023. Trimbos-instituut, Utrecht & WODC, Den Haag.

[12] Netherlands Institute of Mental Health and Addiction (2022). Nationale Drug Monitor: Annual report 2021.

[13] European Monitoring Centre for Drugs and Drug Addiction (2022), European Drug Report 2022: Trends and Developments, Publications Office of the European Union, Luxembourg.

[14] European Monitoring Centre for Drugs and Drug Addiction (2022), New psychoactive substances: 25 years of early warning and response in Europe. An update from the EU Early Warning System (June 2022), Publications Office of the European Union, Luxembourg.

[15] European Monitoring Centre for Drugs and Drug Addiction (2020), New psychoactive substances: global markets, glocal threats and the COVID-19 pandemic. An update from the EU Early Warning System (December 2020), Publications Office of the European Union, Luxembourg.

[16] European Monitoring Centre for Drugs and Drug Addiction (2005-2018), EMCDDA - Europol Annual Report on the Implementation of Council Decision 2005/387/JHA. Retrieved from https://www.emcdda.europa.eu/publications-seriestype/implementation-reports_en

[17] DIMS (2022). 30 years of drugs monitoring in the Netherlands. Retrieved from https://www.trimbos.nl/kennis/feiten-cijfers-drugshandel-drugsincidenten/monitoring-drugsmarkt-dims/30-jaar-dims/geschiedenis-van-het-dims/

[18] Miliano, C., Margiani, G., Fattore, L., & De Luca, M. A. (2018). Sales and advertising channels of new psychoactive substances (NPS): internet, social networks, and smartphone apps. Brain sciences, 8(7), 123.

[19] Davey, Z.; Schifano, F.; Corazza, O.; Deluca, P. Psychonaut Web Mapping Group e-Psychonauts: Conducting research in online drug forum communities. J. Ment. Health 2012, 21, 386–394.

[20] Ornella Corazza, Giuseppe Valeriani, Francesco Saverio Bersani, John Corkery, Giovanni Martinotti, Giuseppe Bersani & Fabrizio Schifano (2014) "Spice," "Kryptonite," "Black Mamba": An Overview of Brand Names and Marketing Strategies of Novel Psychoactive Substances on the Web, Journal of Psychoactive Drugs, 46:4, 287-294, DOI: 10.1080/02791072.2014.944291

[21] Rosa Koenraadt & Katinka van de Ven (2018). The Internet and lifestyle drugs: an analysis of demographic characteristics, methods, and motives of online purchasers of illicit lifestyle drugs in the Netherlands, Drugs: Education, Prevention and Policy, 25:4, 345-355, DOI: 10.1080/09687637.2017.1369936

[22] Evans-Brown, M.J., McVeigh, J., Perkins, C., & Bellis, M.A. (2012). Human enhancement drugs. The emerging challenges to public health. Liverpool: Liverpool John Moores University

[23] Barenholtz, E., Krotulski, A. J., Morris, P., Fitzgerald, N. D., Le, A., Papsun, D. M., & Palamar, J. J. (2021). Online surveillance of novel psychoactive substances (NPS): Monitoring Reddit discussions as a predictor of increased NPS-related exposures. International Journal of Drug Policy, 98, 103393.

[24] Rhumorbarbe, D., Morelato, M., Staehli, L., Roux, C., Jaquet-Chiffelle, D. O., Rossy, Q., & Esseiva, P. (2019). Monitoring new psychoactive substances: Exploring the contribution of an online discussion forum. International Journal of Drug Policy,

73, 273-280.

[25] Goossens, F. X., Frijns, T., Van Hasselt, N. E., & Van Laar, M. W. (2013). Het grote uitgaansonderzoek 2013. Uitgaanspatronen, middelengebruik en risicogedrag onder uitgaande jongeren en jongvolwassenen. Utrecht: Trimbos-instituut.

[26] Monshouwer, K., Van der Pol, P., Drost, Y. C., Van Laar, M. W. (2016). Het grote uitgaansonderzoek 2016. Utrecht: Trimbos-instituut.

[27] K. Monshouwer, C. van Miltenburg, R. van Beek, W. den Hollander, F. Schouten, M. van Goor, M. Blankers, M. van Laar (2020). Het grote uitgaansonderzoek 2020. Utrecht: Trimbos-instituut.

[28] Barratt, M. J., Lamy, F. R., Engel, L., Davies, E., Puljevic, C., Ferris, J. A., Winstock, A. R. (2022). Exploring Televend, an innovative combination of cryptomarket and messaging app technologies for trading prohibited drugs. Drug and Alcohol Dependence, 231, 109243. https://doi.org/10.1016/j.drugalcdep.2021.109243

[29] Blankers, M., van der Gouwe, D., Stegemann, L., Smit-Rigter, L. (2021). Changes in Online Psychoactive Substance Trade via Telegram during the COVID-19 Pandemic. European Addiction Research, 27(6), 469–474. https://doi.org/10.1159/000516853

[30] Bosman, S. (n.d.). How criminals are tracked down on Telegram. Retrieved January 20, 2023, from https://www.cybersprint.com/news/how-criminals-are-tracked-down-on-telegram

[31] No-SIM Signup, Auto-Delete All Chats, Topics 2.0 and More. (2022, December 6). Telegram. https://telegram.org/blog/ultimate-privacy-topics-2-0?setln=en

[32] Nielsen, J. (2000, Mar 18). Why You Only Need to Test with 5 Users. Nielsen Norman Group. Retrieved February 7, 2023, from https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

[33] Nielsen, J. (1994, Nov 1). Severity Ratings for Usability Problems: Article by Jakob Nielsen. Nielsen Norman Group. Retrieved February 7, 2023, from https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/

7 APPENDIX

7.1 Relevant sub-Reddits

r/1P_LSD, r/2cb, r/4acodmt, r/adderall, r/afinil, r/ambien, r/AMT, r/Ayahuasca, r/cocaine, r/cripplingalcoholism, r/DMT, r/DPH, r/dxm, r/fentanyl, r/ketamine, r/kratom, r/LSA, r/LSD, r/MDMA, r/MemantineHCl, r/mescaline, r/meth, r/PCP, r/phenibut, r/PsilocybinMushrooms, r/Salvia, r/shroomers, r/shrooms, r/benzodiazepines, r/dissociatives, r/DissonautUniverse, r/gabagoodness, r/microdosing, r/noids, r/opiates, r/_RCs, r/PsychedelicMessages, r/Psychedelics, r/PsychedSubstance, r/Psychonaut, r/RationalPsychonaut, r/Stims, r/treedibles, r/trees, r/tryptonaut, r/AnAnswerToHeal, r/AskDrugNerds, r/askdrugs, r/aves, r/bestoferowid, r/CurrentlyTripping, r/druganalytics, r/druggardening, r/DrugNerds, r/Drugs, r/DrugShowerThoughts, r/DrugsOver30, r/erowid, r/ObscureDrugs, r/ReagentTesting, r/researchchemicals, r/samelevel, r/TheDrugClassroom, r/TheeHive, r/tripreports, r/TripSit, r/TripTales.

7.2 Final product user validation survey

Dear participant,

We would like to kindly ask you to fill up this survey while you are using our dashboard about New Psychoactive Substances in the Netherlands, monitored through drug forums and Reddit. It helps us to ensure that the dashboard is useful for the end users. It is also a valuable source for improving the design and functionality of our dashboard.

The survey is completely anonymous and, for validation purposes only.

Process:

In this section, you will play the role of a drug investigator and you will be asked to interact with the dashboard choosing drug names to see the number of mentions and views of NPS during a time period.

The dashboard displays lists of our flagging system findings in the top right corner (e.g. top popular NPS, newly occurring NPS, and rising NPS). Additionally, you will find a pie chart of the current percentage of the main NPS views on drugsforum.nl.

On the top left side, the platform contains a linear scale where you can choose the time period of drug mentions and views. Additionally, there is a drop-down search box and a search bar to seek the drug name. You can use either of them to find NPS. After your selection, the dashboard will show line plots under the display settings to see the results. Zooming in and clicking on marked spots is useful to check to see output details.

Questions:

(1) You are doing an investigation about the trends of 3-MMC on drugsforum.nl from 2010 to 2020. Do you find the results?
(2) Now, can you find the results for 2021 only?
(3) You are looking for the number of comments on Reddit for the drug 2-FA on the 1st of September 2022. Can you find it?
(4) You are requested to find the two newly occurring NPS. Please select their names.
(5) During which month there is the highest peak of views in drugsforum.nl for LSD in 2021?
(6) Research about the drug 6-APB is your main project for this year and you are requested to find the percentage of views of 6-APB over the total number of views in Drugsforum.nl. Please write down its percentage number.
(7) You are investigating the most popular NPS to finish a police report. Please select their names from the following options.
(8) Please feel free to give further comments or to give your feedback. Thank you for your time!

7.3 NPS list from vendor webshops

187 NPS identified. Last update: January 18, 2023.

1B-LSD	4F-ADB	DICLAZEPAM
1D-LSD	4F-AMB	ED-DB
1P-LSD	4F-MAR	EPT
1V-LSD	4F-MDMB-2201	ETH-LAD
1CP-AL-LAD	4F-MPH	ETHYL-HEX
1CP-LSD	4F-METHYLFENIDAAT	ETHYL-PENTEDRONE
1CP-MIPLA	4F-RITALIN	ETIZOLAM
2+3-FEA	5-APB	FLUALPRAZOLAM
2-CB	5-BR-DMT	FLUBROMAZEPAM
2-FA	5-BROMO-DMT	FLUBROMAZOLAM
2-FEA	5-CL-ADB-A	FLUBROTIZOLAM
2-FMA	5-DBFPV	FLUETIZOLAM
2-FPM	5-EAPB	FLUNITRAZOLAM
2-MMC	5-HTP	FLUOREXETAMINE
2C-B-FLY	5-MAPB	GW-0742
2C-C	5-MEO-DALT	GW-501516
2C-D	5-MEO-DIPT	GIDAZEPAM
2C-E	5-MEO-MIPT	HEP
2C-E-NBOME	5-MMPA	HEX-EN
2F-DCK	5-MEO-DMT	HXE
2F-KETAMINE	5-MEO-MET	IDRA-21
3-CEC	5-METHYLETHYLONE	JWH-210
3-CMC	5BR-ADB-INACA	L-THEANINE
3-CPM	5F-ADB	LGD-4033
3-FA	5F-MDMB-2201	LSD
3-FEA	5F-PCN	LSZ
3-FMA	5F-SGT-151	MD-PHP
3-FPM	6-APB	MDPHP
3-HO-PCE	6-CL-ADB-A	MEAI
3-HO-PCP	6-CL-ADBA	MET
3-MEC	7-ABF	MF-PVP
3-MEO-PCE	7-ADD	MK-2866
3-MMA	A-PCYP	MK-677
3-MMC	A-PHP	MXPR
3-ME-PCE	A-PIHP	MXIPR
3-ME-PCP	ADB	MEPHEDRENE
3-ME-PCPY	ADB-BUTINACA	METHALLYLESCALINE
3-MEO-PCP	AL-LAD	N-ETHYL-HEXEDRONE
3.4-DMMC	ALD-52	N-ETHYLHEXEDRONE
3D-MXE	A-D2PV	NB-5-MEO-DALT
4-ACO-MET	AMT	NB-5-MEO-MIPT
4-ACO-DET	-PIHP	NDH
4-ACO-DPT	BB-8	NEP
4-ACO-MIPT	BK-2C-B	NORFLURAZEPAM
4-CDC	BK-BBDP	O-DSMT
4-CEC	BK-EBDP	O-PCE
4-CL-PVP	BK-MDDMA	PHENIBUT
4-CMC	BOH-2C-B	PHOSPHATIDYLSERINE
4-EMC	BROMAZOLAM	PYRAZOLAM
4-FMA	BROMONORDIAZEPAM	RAD-140
4-FMPH	CBD	RAD-150
4-HO-DET	CLONAZOLAM	RIBOFLAVINE
4-HO-DPT	DC-C	S-23
4-HO-EPT	DC-TROPA-MIX	S-4
4-HO-MALT	DCK	SGT-152
4-HO-MET	DESOXY-MDA	SR-9009
4-HO-MIPT	DESOXY-MDMA	SR-9011
4-HO-MCPT	DHM	SULBUTIAMINE
4-ME-MABP	DMC	SYNTHACAINE
4-MEC	DMXE	THC-C4
4-MPD	DOC	TRYPTAMINE
4-MPM	DPT	VINPOCETINE
4B-MAR	DESCHCLOROKETAMINE	YK-11
4C-MAR	DESCHLOROETIZOLAM	

7.4 Results correlation table

drugsforum.nl																		
NPS	Views						Comments						Mentions					
	Actual	p-val	Average	p-val	PLR	p-val	Actual	p-val	Average	p-val	PLR	p-val	Actual	p-val	Average	p-val	PLR	p-val
2C-B	0.53	0.1153	0.58	0.0816	0.56	0.0978	0.78	0.0075	0.78	0.0079	0.79	0.0083	0.78	0.0072	0.78	0.0083	0.78	0.0064
3-MMC	0.94	0.0	0.92	0.0002	0.92	0.0002	0.92	0.0002	0.88	0.0007	0.88	0.0008	0.94	0.0001	0.92	0.0002	0.92	0.0002
4-MMC	0.67	0.0325	0.69	0.026	0.69	0.0318	0.54	0.1068	0.56	0.0924	0.60	0.0912	0.54	0.1106	0.51	0.1321	0.49	0.107
4-FA	0.76	0.0101	0.82	0.0038	0.82	0.0025	0.60	0.0644	0.65	0.0414	0.65	0.0388	0.66	0.0384	0.73	0.0165	0.73	0.017
4-FMA	0.50	0.139	0.49	0.1498	0.47	0.1777	0.18	0.6094	0.16	0.6633	0.15	0.7366	0.43	0.21	0.40	0.253	0.39	0.2508
Average	0.68		0.70		0.69		0.60		0.61		0.61		0.67		0.67		0.66	
drugsforum.info																		
NPS	Views						Comments						Mentions					
	Actual	p-val	Average	p-val	PLR	p-val	Actual	p-val	Average	p-val	PLR	p-val	Actual	p-val	Average	p-val	PLR	p-val
2C-B	-0.76	0.0106	-0.76	0.0114	-0.75	0.0119	-0.69	0.0286	-0.71	0.0207	-0.7	0.0225	-0.84	0.0024	-0.82	0.0037	-0.81	0.0043
3-MMC	-0.63	0.0504	-0.63	0.0486	-0.63	0.0504	-0.47	0.1729	-0.47	0.1674	-0.48	0.1623	-0.14	0.6974	-0.16	0.6605	-0.16	0.6652
4-MMC	-0.82	0.0034	-0.078	0.0073	-0.79	0.0066	-0.8	0.0056	-0.83	0.0029	-0.87	0.0012	-0.77	0.0087	-0.77	0.0085	-0.78	0.0076
4-FA	-0.33	0.3573	-0.31	0.3841	-0.31	0.3828	-0.27	0.4479	-0.25	0.4827	-0.28	0.4264	0.16	0.6555	0.23	0.5167	0.25	0.4805
4-FMA	-0.2	0.5729	-0.21	0.5669	-0.21	0.5675	-0.19	0.6009	-0.21	0.5622	-0.21	0.5608	-0.13	0.7238	-0.16	0.6535	-0.17	0.6325
Average	-0.55		-0.40		-0.54		-0.48		-0.49		-0.51		-0.34		-0.34		-0.33	
reddit																		
NPS	Upvotes						Comments						Mentions					
	Actual	p-val	Average	p-val	PLR	p-val	Actual	p-val	Average	p-val	PLR	p-val	Actual	p-val	Average	p-val	PLR	p-val
2C-B	0.66	0.0373	0.72	0.0199	0.72	0.020	0.66	0.0365	0.75	0.0131	0.75	0.013	0.79	0.0065	0.82	0.0035	0.82	0.004
3-MMC	0.88	0.0007	0.88	0.0008	0.88	0.001	0.86	0.0013	0.84	0.0023	0.84	0.002	0.89	0.0006	0.87	0.0011	0.87	0.001
4-MMC	0.56	0.0944	0.57	0.0875	0.56	0.089	0.56	0.0931	0.66	0.0371	0.66	0.037	0.69	0.0263	0.72	0.0192	0.71	0.021
4-FMA	0.14	0.6935	0.14	0.6935	0.14	0.693	0.14	0.6935	0.14	0.6935	0.14	0.693	0.14	0.6935	0.14	0.6935	0.14	0.693
4-FA	0.05	0.9006	-0.26	0.4746	-0.25	0.489	0.13	0.7175	-0.21	0.5549	-0.21	0.563	0.15	0.6794	-0.19	0.5967	-0.19	0.606
Average	0.46		0.41		0.41		0.47		0.44		0.43		0.53		0.47		0.48	

Table 7. Correlation between different yearly popularity metrics from drugsforum.nl, drugsforum.info and Reddit, with Trimbos test sample data. Here, 'Average' refers to rolling average of the actual data, and 'PLR' is the data obtained after applying piecewise linear regression on the rolling average. All values > 0.70 (high positive correlation) are indicated in red.
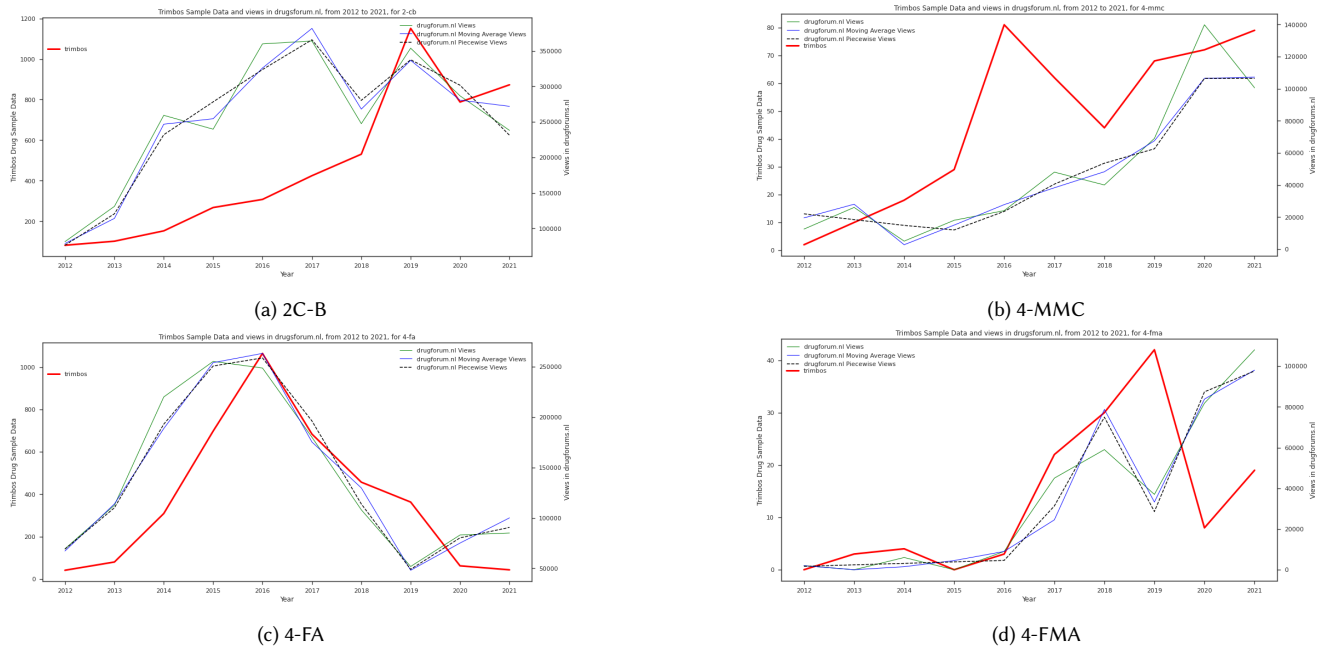
7.5 Results plots of drugsforum.nl and Reddit

(a) 2C-B

(b) 4-MMC

(c) 4-FA

(d) 4-FMA

Fig. 3. Yearly summed plots for number of views on drugsforum.nl of 2C-B, 4-MMC (Mephedrone), 4-FA (4-FMP) and 4-FMA.
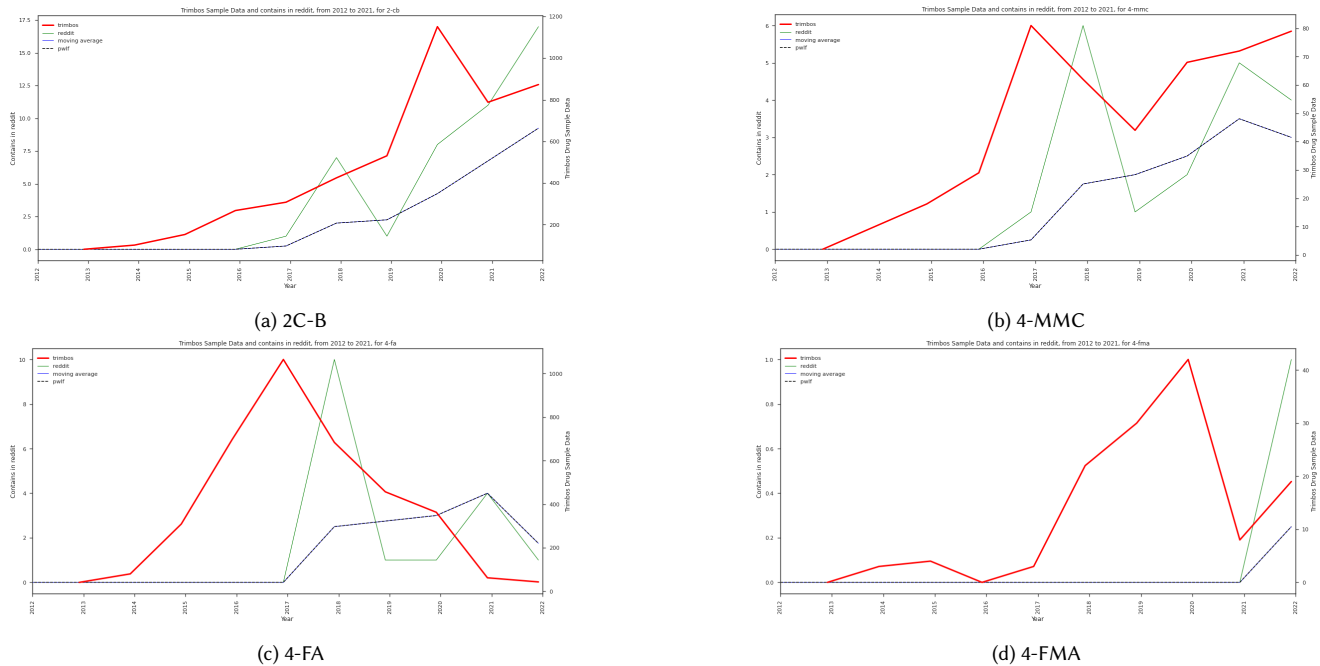
(a) 2C-B

(b) 4-MMC

(c) 4-FA

(d) 4-FMA

Fig. 4. Yearly summed plots for number of mentions on Reddit of 2C-B, 4-MMC (Mephedrone), 4-FA (4-FMP) and 4-FMA.

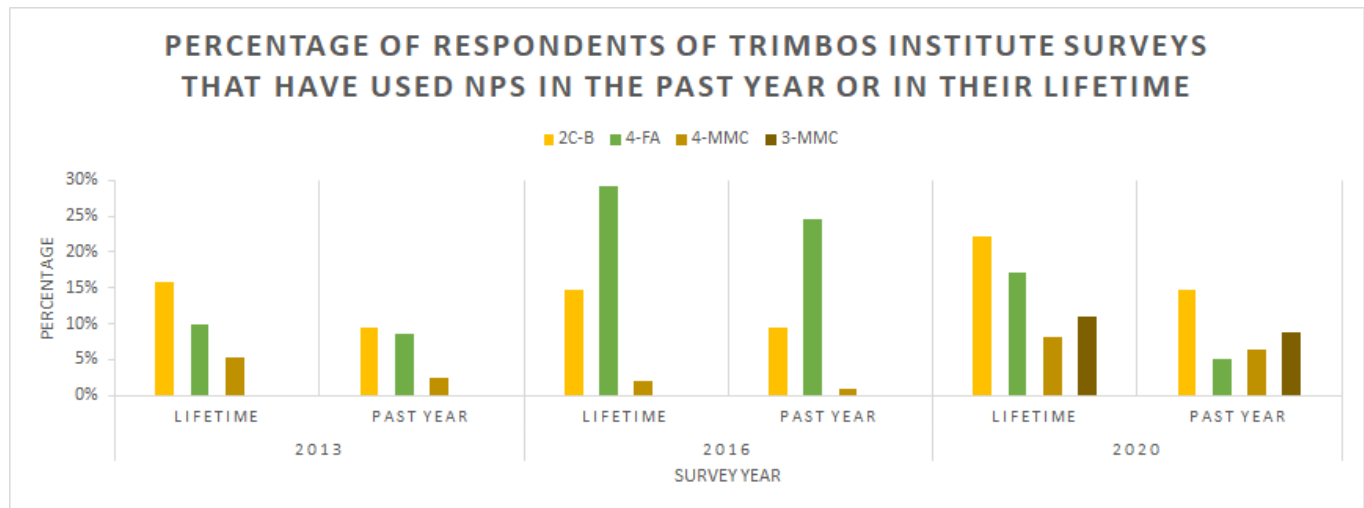## 7.6 Trimbos institute NPS popularity data

Fig. 5. Survey data on NPS use of *Het grote uitgaansonderzoek* from Trimbos Institute [25–27].