**Take-aways:**

**Most of the papers focus on 1 source of information (drug-forum, reddit etc.) to find trends in popularity, not several.**

**Select NPS on which data is available, so not the newest. We need to validate our findings.**

**Problem is universal in all papers -> rapid** **emergence of NPS -> goal is to use online data from social media/forums as a additional source for an early warning system -> good results, often popularity on social media increases before the EWS flags a new NPS**

**Beautifulsoup 4.4.0 used for scraping Drug-forum**

**Regular expressions often used**

**Article 1: Online surveillance of novel psychoactive substances (NPS): Monitoring Reddit discussions as a predictor of increased NPS-related exposures**

**Problem** -> NPS account for an increasing proportion of adverse events, hospitalizations, and deaths due to increasing potency and unanticipated biological effects compared to predecessors.

**Goal** -> utility of drug use forums as an early indicator or predictor of impending intoxications with potentially harmful or lethal outcomes prior to their occurrence

**Data sources** -> Reddit + data from NPS Discovery at the CFSRE and NMS Labs (open-access analytical drug early warning system in the US that specializes in NPS test-

ing, data compilation, and information dissemination)

**Method** ->

1. 8 NPS were selected due to the knowledge that mentions existed on Reddit and toxicology cases were previously reported, providing an appropriate dataset for analysis. In addition, all 8 NPS have received some level of notoriety due to prevalent involvements in adverse events and/or death. -> *carfen- tanil (a synthetic opioid), U-47700 (a synthetic opioid), eutylone (a synthetic stimulant/cathinone), flualprazolam (a new benzodiazepine), N -ethylpentylone (a synthetic stimulant/cathinone), 5F-MDMB-PICA (a synthetic cannabinoid), isotonitazene (a synthetic opioid), and brorphin*e
2. Reddit was selected as the drug forum of choice for this study due to prior anecdotal experience with qualitatively monitoring drug user experience self-reporting on the site and the large user base. Online data consisted of raw counts of keywords related to the eight selected NPS that appeared on Reddit forums from 2010 to 2020, spanning one decade. We first identified specific subreddits that were likely to contain discussions of NPS, using two criteria for inclusion: 1) subreddits that contain frequent mentions of drug terms that were designated as

relevant to our analysis (e.g., /r/Opioid\_RCs) and 2) subreddits that self-identified, by title or description, as dedicated to drug discussion. Among those self-identified subreddits, we distinguished three basic categories: 1) drug specific (e.g., /r/kratom, /r/cocaine), 2) dedicated to a drug class (e.g., /r/Stims, /r/opiates), and 3) general psychoactive substances (e.g., /r/researchchemicals, /r/DrugsOver30). Based on these criteria, 67 subreddits were identified for inclusion.

1. To capture a reliable measure of a drug’s presence in Reddit discussions, we selected and searched only for keyword terms that refer to the specific substance directly. These keywords can include drug names, abbreviations, and slang terms. In order to identify substance- specific keywords, we started with standard terms from drug nomenclature. Next, we used an unsupervised algorithm for keyword detection to search subreddit comments and identify additional terms referring to the same substance. The unsupervised method comprised an application of the Term Frequency-Inverse Document Frequency (TF-IDF) technique, which counted the frequency of every word in every comment to find terms that appear more often in a drug subreddit than the average subreddit ( Jones, 2004 ; Luhn, 1957 ). Terms frequently used in drug subreddits were assessed manually to determine whether they were being used as alternative terms for a substance. To preserve keyword specificity, a term was only included if its mentions could be unambiguously attributed to a substance when used in drug subreddits. Keyword mentions were counted for each drug by searching all comments in every selected subreddit. Mentions were counted based on regular expression matching (non-case sensitive) of drug keywords to substrings of comment text. Matches were aggregated to a set of comments, each identified by a unique identifier in Reddit’s API. Thus, a comment was only counted once if the same substance is mentioned multiple times in the single comment’s text. Comments that were counted within thesame unit of time, such as a calendar month or quarter, were summed to a single count. Counts were compared over time to reveal trends in online drug mentions.
2. Data from the analysis of toxicological samples for the presence of drugs, including the targeted NPS, were compiled from NPS Discovery at the CFSRE (Willow Grove, PA, USA) and NMS Labs (Horsham, PA,USA). NPS Discovery [ www.npsdiscovery.org ] is an open-access analytical drug early warning system in the US that specializes in NPS testing, data compilation, and information dissemination. largest forensic toxicology reference laboratory in the US providing national testing services in forensic toxicology and drug chemistry laboratories, with an approximate average yearly case load of more than 120,000 cases. Data were generated from comprehensive toxicological analyses performed on biological specimens submitted in suspected death and impaired driving investigations. Following analytical con- firmation, drug testing results, consisting of the identity of the drugs present and the date of analysis, were output to an in-house data lake (a large repository of raw data elements). This data lake was queried to produce data tables containing the frequency of detection of specific drugs over time. The aggregate data were tallied and tabulated by year and month. First incidence of detection and identification from NPS Discovery were combined with data from NMS Labs to generate a comprehensive timeline of frequency of detection.
3. Analysis: The number of Reddit mentions and number of related exposures were aggregated by month and quarter within each year as described above. First, data were plotted using observed values (raw counts) by month, and then trends were estimated using quarterly data. Quarterly data were used to estimate trends due to monthly data often containing multiple time points with very few cases. Joinpoint Regres- sion (Version 4.8.0.1) was used to evaluate trends ( NCI, 2020 ). Also known as piecewise, broken line, multi-phase, or segmented regression, Joinpoint fits weighted least-square regressions to rates on a log transformed scale ( Ingram et al., 2018 ; Kim, Fay, Feuer, & Midthune, 2000 ). This program also uses Monte Carlo permutation tests with a Bonferroni correction for multiple testing and identifies the best- fitting set of joinpoints within models. Joinpoints are knots in trends that indicate significant shifting points. Separate joinpoint analyses ( n = 16) were conducted on Reddit mentions and exposures for all eight NPS.

**Results** -> Seven NPS (excluding 5F-MDMB-PICA) appeared in discussions on Reddit prior to their implication in poisonings or intoxications. Distinct increases and decreases in number of mentions and number of exposures were observed. For most substances ( n = 5, 63%), a rise in Reddit mentions was soon followed by a corresponding rise in toxicology positivity. Peak positivity for carfentanil and flualprazolam, however, preceded peak Reddit mentions. For seven out of the eight NPS assessed, online mentions preceded observed exposures with gaps between 8 months and > 4 years. The timeframe (mean = 5.2 months, median 5 months, range 1-11 months, n=5) between peak mentions and peak exposures varied.

Afbeelding met tekst

Automatisch gegenereerde beschrijving

**Conclusion** -> This study demonstrated the utility of social media sites, such as Reddit, as a predictor for future trends in NPS-related exposures. These results provide confirmation that activity on drug use forums in the virtual world can help predict changes in exposures associated with new or re-emerging NPS in the real world. The results warrant further evaluation as a strategy for inclusion in early warning system.

**Article 2: Monitoring new psychoactive substances: Exploring the contribution of an**

**online discussion forum**

**Problem** -> The rapid emergence of new psychoactive substances (NPS) is a challenge for public health authorities and law enforcement. A modification of the substance structure can easily be done

once the legislation to prohibit this substance is passed (Chatwin, Measham, O’Brien, & Sumnall, 2017). The phenomenon is strengthened since the increase of the Internet usage.

**Goal** -> This research aims at understanding the contribution of data extracted from a major online discussion forum within a systematic monitoring process. In the context of NPS, monitoring substances through the analysis of discussion forums could help identify the most discussed substances among consumers. Such a result could be used as a proxy to assess the popularity of substances over time. It could help target efforts (i.e. law enforcement as well as prevention) on specific problematic substances and help policy makers and practitioners to make progress in responding to this “new” phenomenon. As such, it would help strength the already in place early-warning systems.

**Approach** ->

1. Drugs-forum opened in 2003 and now includes more than 250,000 registered members and a wide range of discussions (i.e. more than 220,000 discussion threads). This forum was chosen because of its size.
2. The subsections of this forum listed in the section “Drug information & harm reduction” were considered for this study. A Python-coded system was designed to crawl the Drugs-Forum website and scrape the data (versions: Python 2.7 and its library BeautifulSoup 4.4.0). Briefly, this system takes as an input the sections’ main URL (i.e. web address) and builds a list of its respective following pages. All these pages are then browsed, data related to each discussion thread is duplicated (restructured if necessary) and stored into a dedicated database. Data related to 84,678 discussion threads was collected. In particular, the title of each thread was extracted, as well as its date of creation, and the number of views and replies it generated. The number of users who took part in the discussions was also recorded. Further analysis was conducted using the Tableau Desktop 10.5 visualisation software.
3. The United Nations Office on Drugs and Crime (UNODC) categorises NPS in nine classes namely phenethylamines, synthetic cannabinoids, synthetic cathinones, piperazines, tryptamines, aminoindanes, phencyclidine- type substances, plant-based substances, and other substances. A selection of the most common NPS reported to the UNODC Early Warning Advisory on NPS was made to cover each class except for the other class.

----

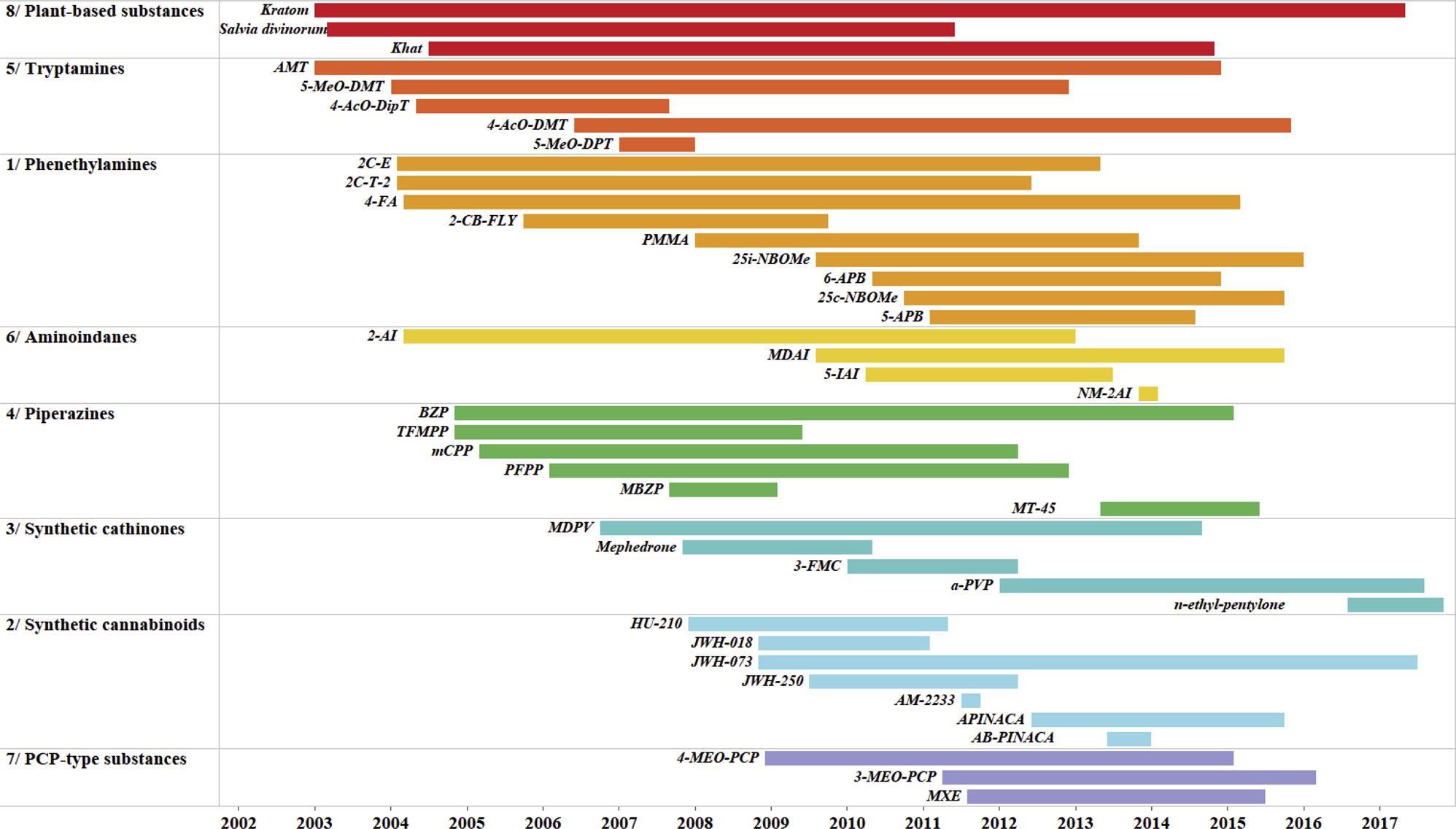
1. The indicators available to assess the popularity of a substance were the number of views, the number of replies, the number of distinct authors involved in a discussion thread, and the number of threads related to this substance (Pineau et al., 2016). Regular expressions were used to search for the name of substances in the titles while taking into account the variations of a substance name (e.g. JWH-018 might also be written ‘JWH 018’, ‘Jwh018’, etc). A matrix of Spearman’s rank correlation coefficients was calculated between each of the indicators. Each calculated correlation was above 0.93 (p-value<0.001). Thus, any indicator may be chosen to assess the popularity of a substance -> the cumulative number of replies across all threads related to a substance was selected to assess its popularity. The other indicators were not chosen since they might not reflect users’ interest in a substance.
2. Analysis: First, the timeframe during which a substance was present (i.e. discussed) on the discussion forum was calculated and represented over time. Secondly, the changes in popularity of this substance over time were studied in more details. Two conditions were required to consider a substance as being present on Drugs-forum during a given month. First, threads related to the substance should have replies over the month. Second, the monthly number of replies related to this substance represented at least 1% of the total number of replies across the threads related to this substance. The timeframe during which a substance was present on the forum was then visualised. This timeframe starts with the creation date of the first thread discussing the substance. It ends with the creation date of the last thread that generated at least 1% of the replies, without considering any interruption. Changes in popularity for each substance was investigated over the timeframe. As performed by Pineau et al. (2016) in the context of doping products, the number of replies was normalised and represented over time. Different patterns were then highlighted visually (i.e. chronic, downward, upward), as well as particular peaks of interest.

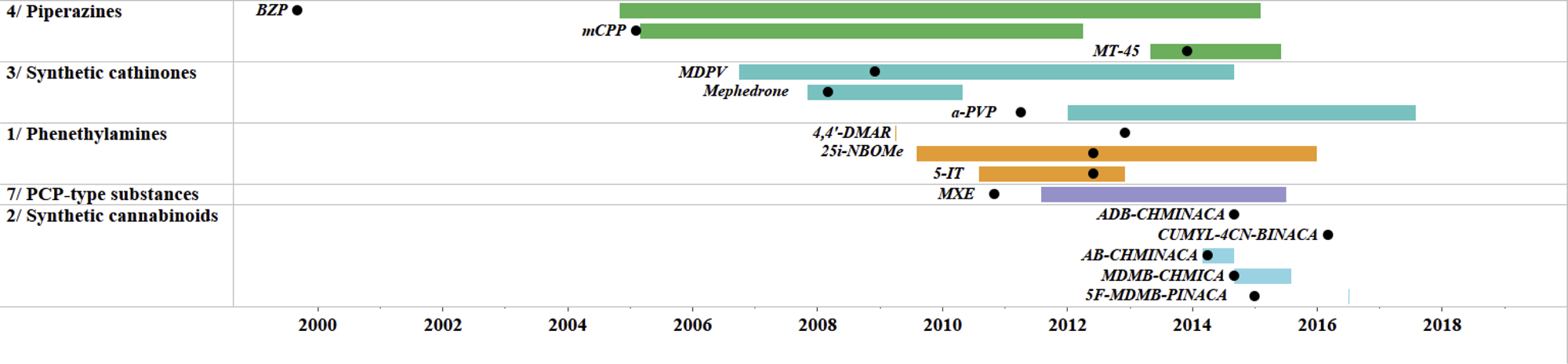
**Results** -> Out of the fifteen substances subjected to a risk assessment from the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), nine of them on the forum before or at the same time as their first date of notification to the EU early warning system. The trends drawn by analysing this online discussion forum reflect the information from the literature. This tends to confirm that such forums are a valid source to monitor the appearance of substances on the market. Generally, the popularity of most substances on the forum follows a downward trend. Several hypotheses can be raised to explain this ob

servation. It could indicate a loss of interest for the substance, due to a

decrease in availability or negative health effects reported. Alternatively, the early discussions could have covered a large number of topics, implying that potential consumers are able to find answers to

their questions without having to write replies or create new threads.





**Conclusion** -> In line with previous research on doping products, this article shows the potential of the monitoring of online forums in the context of psychoactive substances. Besides, the system designed to collect the data is flexible and can be systematically updated to fuel a monitoring process. It informs not only on the presence/absence of a substance in discussions between consumers, but also on its evolution over time. Such results could benefit academic research and organizations studying the NPS phenomenon. Precisely, it could complement existing early warning systems and benefit law enforcement agencies and policy makers. In order to strengthen early-warning systems, alerts could also be defined based on sudden

rises or temporary spikes, whether the spike is due to positive or negative discussions. The former could inform on the possibility to see this substance more frequently on the market, while the later could trigger

preventives measures.

**Article 3: Analysis of Google Trends to monitor new psychoactive substance. Is there an added value?**

**Problem ->**

NPS are challenging for public health authorities, therefore the two major drug monitoring bodies – the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) and the United Nations Office on Drugs and Crime (UNODC) have implemented the EU Early Warning System (EWS) and Early Warning Advisory (EWA), respectively. While these monitoring systems are informative, it is difficult to keep up with the constant and rapid developmental rate of NPS. The EMCDDA has recognised the need for an alternative and technologically derived early warning system.

**Aim ->**

The aim of this research is to determine whether Google Trends and drug discussion forum data can be used to complement early warning systems for NPS.

**Approach ->**

Forty-eight substances were used in this study and classed into groups based on their chemical structure, following the UNODC classification system. Google Trends data (time range: 2004–2019) and drug forum data (time range: 2003–2018) were extracted for each substance and visual trend profiles were created for class groups as well as individual substances. Analysis was conducted to determine when a substance first appeared on Google Trends and a drug discussion forum as well as their trends over time. This date of first appearance was then compared to the date the substance was first reported to UNODC.

1. A total of forty-eight distinct NPS were selected as target substances (see Table 1). This selection was based on the most common NPS reported to the UNODC Early Warning Advisory on NPS between 2009 and 2015 and covered the main classes of NPS of the UNODC chemical classification system (i.e. structural similarity to a parent compound) [33].
2. On March 16th, 2019, a Google Trends search was conducted for each of the forty-eight distinct NPS terms outlined in Table 1. For each term, the following parameters were specified: (1) geographical location: Worldwide, (2), time range: 2004 –2019 (i.e. 01/01/2004 – 16/03/2019), (3) category: All categories. Data (i.e. search volume) was collected through the exportation tool available on the website. The Google Trends output is an index of the relative search volume; the data is scaled according to the average search quantity for the term (over the selected time period and geographical location), and normalised to provide a relative volume where maximum search interest is ‘100’ and minimum search interest is ‘0’ [22,35–37].
3. Moreover, whilst the website contains a number of sub-sections, only one entitled ‘Drug Information & Harm Reduction’ was used for this study. A Python-coded system (Python 2.7 and BeautifulSoup 4.4.0) was designed to crawl the section’s data and scrape the needed information from 2003 to 2018 [20]. The system works in two stages: (1) identification of discussion threads related to selected substances and (2) extraction of the number of replies and views and of the time of all subsequent discussion threads. The number of replies normalised over the total number of replies for each substance during the period investigated was used as an indicator of discussion intensity for each substance [20]. In order to obtain the same scale as the Google Trends data, the normalised replies were provided as a percentage (from 0 to 100).
4. The UNODC has recorded reports of the presence of NPS since the beginning of 2009 via the Global Synthetics Monitoring: Analyses, Reporting and Trends (SMART) Programme, but as of 2013 this is reported by the Early Warning Advisory on NPS (EWA) [16,39]. The presence of NPS can be reported to the EWA by a variety of sources, including government authorities, laboratories, and partner organisations [40]. When the EWA is informed of the presence of a previously unreported NPS, the year of first reporting is recorded.
5. Data from Google Trends and Drugs-Forum.com were combined and the start and end dates of appearance of each NPS were recorded. Google Trends start dates were identified as starting on the first date a sufficient number of searches were conducted to producea non-zero result in the dataset. Drugs-Forum.com start dates were identified as starting at the creation date of the first thread discussing that substance. Changes in the intensity of discussion (used as a proxy for popularity for each substance on both Google Trends and Drugs-Forum) were investigated and compared over time. Trends for each substance can be classified in four groups: upward trends, chronic trends, downward trends, and temporary increase [21]. However, when it comes to analysing trends within a class containing multiple substances, more complex patterns can be detected. The different trend types observed were classified as ‘block’, ‘successive’, and ‘generational’ trends. Microsoft Excel (version 1904) was used for data compilation, pre-treatment and Tableau software (version 2019.1.0) was used for visualisation

**Results ->**

Of the three data sources utilised, substances were most likely to appear on Google Trends first. Amongst the different classes of NPS, discernible trends (‘block’, ‘successive’, and ‘generational’ trends) were observed. These trends reflect the evolution of the manufacture of substances or generations of substances that has been observed in the literature.

The first comparison was made between Google Trends and Drugs-Forum. Six substances were only detected on Google Trends, 31 substances were detected on Google Trends first, 7 on Drugs- Forum first, and 4 simultaneously. For substances appearing first on Drugs-Forum, substances appeared on Google Trends within 1–54 months (median of 7 months). For substances appearing first on Google Trends, substance appeared on Drugs-Forum within 1–88 months (median of 10 months). For substances with enough Drugs-Forum data to detect trends (more than 20 distinct temporal data points derived from at least 1% of the total number of replies across the threads related to a substance), similar trends were observed (see Fig. 1). In some cases, trends vary in regard to the intensity of the discussion, but more often than not, the period of most interest are the same in both data sources. Google Trends gives a more continuous dataset. Indeed, some substances might not be discussed on forums for months.

When comparing solely Google Trends and UNODC, 10 of 21 substances (48%) appeared on Google Trends before the UNODC report year, 3 (14%) were first reported to the UNODC, and 8 (38%) appeared on Google Trends in the same year as their UNODC report. For the substances proactively detected on Google Trends, the difference in reporting year was from 1 to 8 years, with a median delay of 4 years. For substances reported by the UNODC first, the difference in reporting year was either 1 or 5 years, with a median delay of 1 year. Fifteen substances were available across all three sources. Six were not detected on Drugs-Forum. Seven were first recorded on Google Trends, two by the UNODC, two simultaneously by Google Trends and Drugs-Forum, two simultaneously by Google Trends and the UNODC, and two by all sources. Thus, no substances were detected proactively on the forum.

**Conclusion ->**

Google Trends showed the emergence, persistence, or transient nature of substances, which could direct the focus of law enforcement, health organisation and laboratory resources towards a limited number of substances. When one considers the reliance of individual information seeking on the Web as well as the prominence of NPS on the Web, it becomes clear that Google Trends and drug discussion forums could be used as a complement to current early warning systems.

This research allowed us to identify evident succession of substance popularity within classes. Between Google Trends and Drugs-Forum, six of eight classes showed remarkable similarity in substance succession. Substances were more likely to appear on Google Trends first and Google Trends also had a shorter median lag time in appearance of substances compared to Drugs-Forum. The two sources showed significant similarity in their visual trend profiles, with either peaks or falls in popularity aligning or periods of intense discussion centring around the same time.

By combining these two data sources, it was possible to compare them and identify whether substances are increasing or decreasing in popularity in regard to search frequency on The Web.

**Article 4: Online drug monitoring (fact sheet)**

**A web crawler** (also called a spider or spider bot) is software that searches systematically, based on previously specified parameters, one or more websites and indexes or collects the data found. In this way, it provides input for the web scraping process.

**Web scraping** is a process whereby data from internet pages are collected and stored. This can be done manually, but normally when web scraping is discussed it is understood to mean the automated collection of data using software.

**Online monitoring** systematically collects data concerning search behaviour, discussions, user

experiences or drug markets and stores it for further analysis.

Some researchers think that all of the data that can be found on the internet are public and can be freely collected and used. Others think that forum messages should not be freely used, for instance, because forum members do not always realise the possible impact of posting messages on forums. Guidelines have now been formulated for doing this type of research via the internet [10]. To keep on the safe side and to prevent any possible breaches of the privacy legislation, we also decided not to collect any data that could be used to trace an individual forum member or unique pseud

**I-TREND Project**

The objective of this project was to provide help in preventing both health and societal damage due to NPS use and to improve the knowledge about the effects and risks of these substances. The project had various parts: the monitoring of forums where drugs were the subject of communication, the monitoring of web shops that supplied NPS, the carrying out of a survey using an online questionnaire about the use of NPS, and monitoring the composition of NPS that were purchased. The project found that in spite of what is generally advertised in web shops about the purity of the substance on offer ( >99%), in reality the purity of NPS varied considerably. Some NPS were wrongly labelled and some contained chemical analogues (including 25B/C-NBOMe instead of 25I-NBOMe, and pentedrone instead of 3,4-DMMC). But in some cases, the NPS proved to contain completely different substances than what was offered (e.g. pentedrone instead of αMT or 3-FMC instead of 5-MeO-DALT). Moreover, the sales prices differed considerably between the countries. In 2018, we also investigated discussions about new synthetic opioids, including substances related to fentanyl. In 2018, the Trimbos Institute carried out a quick-scan into the benefit and feasibility of monitoring trends in new smoking products via the internet. Here we focused on the iQOS (which stands for I-Quit-Ordinary-Smoking). This is a tobacco product that was introduced in 2014 (in the Netherlands in 2017) in which tobacco is not burnt but heated.

The question at the heart of this quick-scan was if it was possible to use Google Trends to answer the following three questions: (1) is the product popular in the Netherlands when compared to other countries? (2) how has this developed through time? and (3) is there regional variation within the Netherlands in the popularity of the product?

To answer these questions, a link was made between the R statistical environment and Google Trends. In fact, this meant that Google Trends was scraped; using scripting we could consult the Google Trends database. In doing so, we operationalised the popularity as the number of Google searches for the product name.

We requested, analysed and compared the data concerning the number of searches within the Netherlands and in other countries in the last five years.

Every monitoring method has its advantages and disadvantages. Traditional population studies, such as questionnaires, can give a bias when respondents do not want to provide certain information or when the sample does not represent the population being studied.

This factsheet has addressed what online drug monitoring is and how it is used. The experiences that we have gained in recent years of online monitoring have also been discussed as has the added value of online monitoring for the regular substance monitors.