18 November 2024
EMA/348808/2024

Social Media Data for Real World Evidence in Regulatory Decision Making

An expert review report for the HMA/EMA Big Data Steering Group - 2024

Executive Summary

Social media has taken on roles beyond just peer-to-peer communication. Today, social media platforms are used for (mass) communication purposes, but also for entertainment, education as well as to assemble, to lobby, and for marketing purposes to name a few uses. Within the fields of public health and medicine, social media has been used to conduct and share health research, stakeholder communication, share disease and treatment experiences and identify emerging threats.

In the recent years, more and more people globally use social media to find health related information, and increasingly more patients gather on social media platforms to discuss disease related issues, share treatment experiences, give advice or seek support for decision making. Moreover, the COVID-19 pandemic facilitated heightened social media use, partly due to social isolation measures, and partly due to the novelty of the virus and public uncertainty leading people to seek out the fast-paced information available on social media. Consequently, large amounts of patient data are currently present on social media platforms, which could be beneficial for regulatory decision making. In addition, the last 5 years have seen an increase in the number of submissions to the European Medicines Agency (EMA) which leverage social media for various purposes. Therefore, it is crucial for regulators to assess the utility and impact of these data on regulatory decision making to ascertain in which instances these data could be leveraged. In fulfilment of its Big Data vision, the joint HMA/EMA Big Data Steering Group (BDSG) aims at investigating the potential of social media data for regulatory uses through this expert review report.

Various patient data can be extracted from social media, including demographic data, patient experience data (PED), and data related to drug use and disease factors for example. A distinction is made between general social media platforms (e.g. Facebook) and patient-focused social media platforms (e.g. Patients Like Me) where the latter is explicitly focused on health-related topics. A particular feature of social media data is that it is voluntarily generated by patients, health professionals, or other social media users in the real world, outside of a controlled setting.

The current report reviewed 73 relevant scientific articles and reviews identified through a literature search on the Embase database. Findings from these papers were used to formulate a set of regulatory use-cases for social media data, as well as to identify relevant challenges and opportunities related to their use. Ultimately, 7 points for consideration for future actions are proposed to establish value and enable the use of social media data in medicine regulation.

Contents

List of Abbreviations

Abbreviation	Explanation
BDSG	Big Data Steering Group
HMA	Heads of Medicines Agencies
EMA	European Medicines Agency
SGOPED	Spontaneously Generated Online Patient Experience Data
PED	Patient Experience Data
RWD	Real World Data
MAA	Marketing Authorisation Application
HCP	Health Care Professional
NLP	Natural Language Processing
PA	Psoriatic Arthritis
QoL	Quality of Life
ADR	Adverse Drug Reaction
RCT	Randomized Controlled Trial
GDPR	General Data Protection Regulation
API	Application Programming Interface
AI	Artificial Intelligence
ML	Machine Learning
ITF	(EMA's) Innovation Task Force
SA	Scientific Advice
MAH	Marketing Authorisation Holder
IMI	Innovative Medicines Initiative
GLP-1 RA	Glucagon-like Peptide 1 Receptor Agonist
EMRN	European Medicines Regulatory Network
ECDC	European Center for Disease Prevention and Control
EEA	European Economic Area
SMM4H	Social Media Mining for Health Applications

1. Scope and Objectives

This expert review report supports the delivery of the Big Data Steering Group (BDSG) workplan to 2025 (HMA/EMA joint Big Data Steering Group, 2023). To advance the ongoing work on expanding the selection of available RWD sources to support regulatory decision making, the BDSG is looking into how data from social media could be used to inform EU regulatory decision making, and what actions may be proposed for their optimal use. The BDSG was consulted on the report between June-July 2024.

Social media can be defined as online environments, platforms, or applications powered by internet connectivity where users can create content for other users and/or access, interact with and contribute to content generated by other users (Chen & Wang, 2021). They differ from traditional media sources such as magazines or television programs in that content creation is not restricted to certain parties and the level of access to other users' content can depends on the platform as well as regional restrictions and usage coverage (Chen & Wang, 2021).

Social media data refers to the information collected from social media platforms. All social media platforms contain, by definition, content generated by users, e.g. a user created video posted on a social media platform together with the comments and/or responses it elicits, in various formats (text, visual, audio). When created by patients, this type of data is also known as spontaneously generated patient experience data (SGOPED) (Walsh et al., 2022).

Characteristics of social media data are described in further detail in section 6. social media data can capture various types of patient data from demographic information to sentiments and attitudes, and patient experience data (PED) related to quality of life or impact of use of medicines for example. Social media data, when not generated during a clinical trial (CT) is generally considered as RWD. Real world data describe patient characteristics in routine clinical practice and include a wide variety of sources such as data from patient registries, electronic health records, wearable devices, medical claims and data on medicinal product prescriptions as well as data from social media.[1]

[1] Current working definition of RWD (*Reflection paper use real-world data non-interventional studies generate real-world evidence* (European Medicines Agency, 2024))

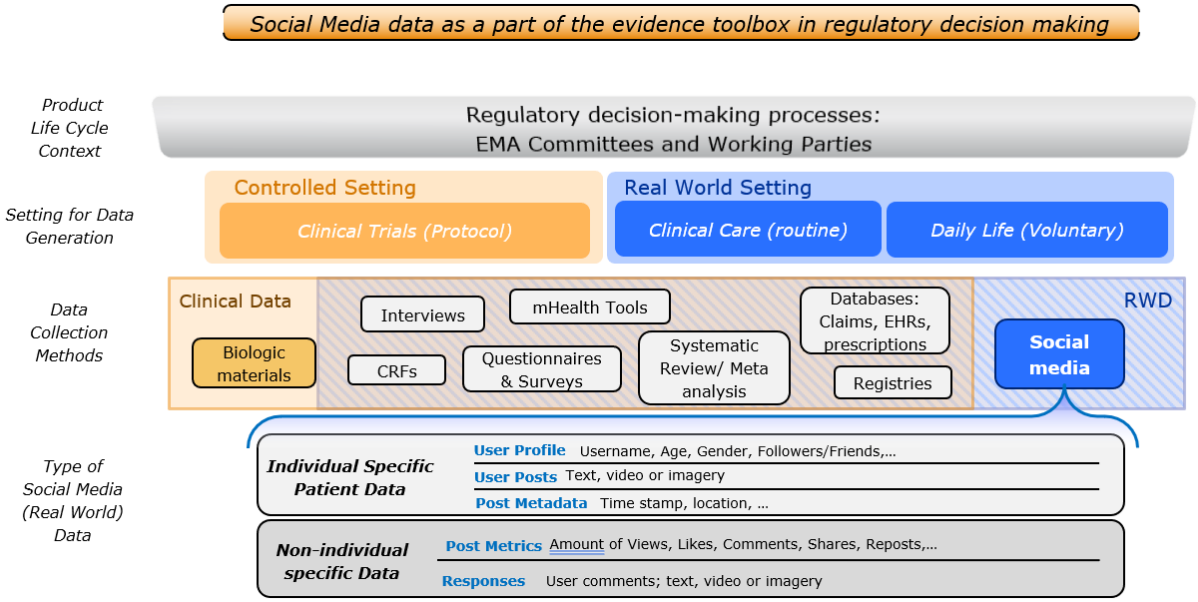


Figure 1. Social Media data in regulatory decision-making. Various regulatory parties make use of data to substantiate their decision-making processes. This data can be generated in either controlled settings through clinical trials, or in a real-world setting during routine care or voluntarily by patients in their daily life. Social media is one of the tools and methods available toto collect patient data. The data generated from social media can be specific to an individual patient, such as their demographic information or text or video content. Or social media data can be non-specific to an individual, such as the metrics present on a user post reflecting the amount of engagement by other users.

Given the diversity of social media platforms and thus of the data available, as well as the variety in possible regulatory use cases, different types of social media platforms are considered in this review. A distinction is made between individual specific data and non-individual specific social media data (Figure 1). A separation is also made between patient focused social media platforms such as Patients Like Me and non-health specific or general social media platforms such as Facebook. This is done for the purpose of addressing the implications a refined content and user base may have on the reliability, relevance, accuracy and ultimately the utility of the data from patient-focused social media platforms compared to the general ones.

This expert review report aims to answer the following questions:

- How can social media data support regulatory decision-making? What are the regulatory use cases and the additional value of social media data?
- Considering the latest technological, methodological, and regulatory developments related to RWD and social media data, what are the current challenges and opportunities for the use of social media data in regulatory decision-making?
- What are the possible points for consideration for future actions to enable the use and establish the value of social media data into the EU regulatory decision-making?

2. Background

The vast leaps in digital technologies over the recent years have allowed for the scientific and regulatory community to look beyond traditional types of data for generating evidence to inform and support regulatory decision-making (Macdonald et al., 2021). More specifically, real world evidence (RWE), generated from the analysis of RWD is increasingly used in regulatory assessments to complement evidence generated through randomized controlled trials (Flynn et al., 2022).

Some types of RWD, like those derived from electronic health records (EHRs) or insurance claims have already been used in marketing authorization applications (MAA), to support the regulatory assessment or for post marketing surveillance purposes (Bakker et al., 2023; European Medicines Agency, 2023; Flynn et al., 2022). Other types of RWD, such as data from social media or mobile health (mHealth), though promising, are yet to be harnessed to their full potential in the regulatory decision-making.

Social media platforms serve many functions for patients in healthcare today, from seeking treatment information and diagnosis to sharing personal experience and offering peer and community support (Oyebode & Orji, 2023). Patients enrolled in overburdened healthcare services often get short and infrequent meetings with their physicians due to high demand and pressure on healthcare services. Therefore, many patients lean onto the internet and online forums to seek health advice and information before going to see a health care professional (HCP) or to supplement the information given during the short and possibly insufficient meetings with their HCP (Oyebode & Orji, 2023). When the COVID-19 pandemic restricted face-to-face social interactions, people also spent more time online and on social media to engage with others on a variety of topics and to voice their opinions (Al-Hamid et al., 2024). Social media became then a channel of heightened importance for HCPs, health care institutions and systems who used innovative remote ways to sustain routine patient care and spread accurate health information while monitoring virus spread via social media platforms (Mohammed et al., 2023). In the recent years social media has taken on a much larger role in advancing health research, offering community support and in the surveillance of disease outbreaks, in large part due to the abundant amount of patient data on these platforms.

This expert review report aims to look at whether both the data and the landscape for use of social media data are sufficiently matured for its use to be further expanded into EU regulatory decision-making.

Since 2020, the joint HMA/EMA BDSG has been pushing the transformation towards an even more data-driven regulatory decision-making. Its Big Data vision imagines a strengthened regulatory system which efficiently integrates the totality of evidence into its assessment processes to support decision-making. Knowing when and how to have confidence in evidence generated through novel technologies will inform and accelerate medicines development, improving treatment outcomes and facilitating earlier and more efficient patient access to new treatments. The 4th workplan of the BDSG published in July 2023 intends to investigate in what capacity social media data could be useful for regulatory decision-making and support the implementation of the European Union Medicines Agencies Network Strategy to 2025 (European Medicines Agency, 2020, 2022a; HMA/EMA joint Big Data Steering Group, 2023).

3. Methods

A literature review was conducted between February 2024 - April 2024 on Embase given its large library of articles focusing on life sciences, medicine and healthcare. Embase was searched without any time restriction, using a mix of keyword searches and subject heading terms to ensure a maximum

coverage of the database. Search terms were developed using the Emtree subject headings. The search terms were applied to the papers' title, abstract, or keywords and the search was filtered for English language publications only. The search terms were refined as the papers were read to include as many possible related themes as possible. Using the final search string, 4640 total search results were obtained, out of which 4091 search results were articles (88.2%) and the rest were reviews (N=549; 11.8%). Articles were included if they analysed use-cases for social media data or discussed relevant challenges and opportunities regarding their use in medicine regulation. Articles were excluded if they did not give relevant information on the utility of social media data for regulatory decision-making. A prioritisations exercise was conducted on the search results, leaving 56 papers included for the literature review. More information on the detailed search strategy can be found in annex 8.1.

Additionally, a preliminary review of internal EMA regulatory pipeline data was undertaken to understand if and through which regulatory procedures is the Agency receiving or being consulted on regarding mHealth data. This review took place in April-May 2024 and at least 18 product related submissions or interactions with the regulators which mentioned use of social media platforms or data were identified between. These submissions were made between 2019-2022. As the review was not exhaustive and all recent and ongoing cases were likely not reflected in the database used, the actual number of submissions is expected to be higher. Findings from the review are further discussed in section 4.

4. Social media in early submissions to the European Medicines Agency

The use of social media data in medicines evaluation is not standard practice yet in EU medicines regulation and remains limited. However, the last 5 years have seen an increase in the number of submissions from industry leveraging social media for various purposes, which is expected to continue increasing in the coming years. A non-exhaustive internal review of an EMA regulatory pipeline database showed that social media has been used or planned for use for patient recruitment, for data collection, to justify applicant study designs, and to evaluate risk minimisation measures (RMMs).

- **Patient recruitment**: Several early submissions through scientific advice (SA) and protocol assistance procedures have described the intent to use social media for recruitment of patients. Both recruitment of patients through targeted social media patient groups as well as broad social media recruitment strategies have been proposed. The Committee for Human Medicinal Products (CHMP) did not comment on the use of social media in most cases, however in some cases expressed concerns over selection bias. A wide range of social media platforms from Facebook to patient focused platforms such as PatientsLikeMe have been proposed.
- **Data collection**: Social media was proposed for the collection of data, mostly via surveys disseminated on social media. Patient-focused social media platforms were proposed for the collection of clinical trial data in a decentralized trial design, while both general and patient-focused social media platforms were suggested for collection of PED or RWD. CHMP did not directly address the choice of social media platforms in most instances, but was supportive to use of social media platforms, with additional attention to the reliability and verification of remotely collected patient data.
- **Study design**: Another capacity in which social media was present in early submissions to the EMA was to inform the choice of study design. Some applicants leveraged previous studies based on social media to describe unmet needs and symptoms for the disease at hand. Other

applicants proposed to use social media to further understand patient experience and impacts on quality of life (QoL) and inform meaningful choices related to study design such as the choice of outcome measurements. Efforts to better understand the patient experience and patient-relevant outcomes were welcomed by the CHMP, though using social media as the medium was generally not commented on.

- **Evaluation of risk minimisation measures (RMMs)**: Social media has also been used to assess the effectiveness of RMMs as a part of the risk management plan (RMP). For example, some applicants have proposed to collect pharmacist and consumer knowledge on proper drug use and safety via social media surveys. In the sample of internal documents analysed the CHMP did not explicitly comment on the choice of using social media for evaluation of RMMs.

5. Use cases for social media data in regulatory decision making

The literature review revealed that use of social media data for routine regulatory decision making has not yet become established. The sections below explore for what use cases social media data could be utilised drawing on recent and relevant scientific literature.

Table 1. Summary table of potential regulatory use cases for social media data.

To Support planning of applicant studies	To inform understanding of clinical context	To investigate impacts and associations	To monitor, prepare for and address public health challenges
Design and Feasibility Recruitment & Screening - Social media allows for simple early assessment of participant eligibility. Data collection - Surveys and questionnaires can be disseminated via social media Patient prevalence - Analysing location information from social media posts makes possible understanding geographic patient and disease prevalence. Study population - Social media allows access to different and often underrepresented populations. E.g. patients with co-morbidities Patient Led Research - Social media may facilitate health research to support drug discovery and medicine regulation. - It particularly increases patient engagement and role in medicine regulation.	***Disease Epidemiology*** Disease spread and epidemiological course - Mentions of disease and symptoms can be tracked, using AI and ML methods for example, to describe spatial and temporal changes in a disease's course. Symptom Clustering - Patient's descriptions on symptoms allow for better understanding of how symptoms cluster in different indications. ***Clinical Management*** Barriers to care - Descriptions of different obstacles restricting patient access to available care is prominent in social media data. Description of clinical care - Patient discussions on care experiences can supplement the understanding of clinical management obtained from other RWD sources. ***Drug Utilisation*** Off Label use, misuse, and abuse - Social media data can be highly advantageous in understanding in what ways medicines are being used beyond their approved indication. Alternate drug use patterns - Analysis of online discussions can reveal different drug schedules and doses patients are following. Unmet needs - Online discussions around current standard of care, e.g. which patients can or cannot access care, which products off- or on-label, patients rely on and how, can help inform areas of unmet medical needs. Reasons for discontinuation - Online discussions can help to describe and understand reasons for why patients discontinue or do not use certain products	***Effectiveness and Safety*** ADR Detection - Monitoring social media allows for cost-effective early detection of ADRs, particularly useful for novel and mild/less severe ADRs. Effectiveness - Standardised questionnaires on effectiveness can be implemented via social media, particularly on patient-focused platforms. - Some insights into drug effectiveness can be made from social media data when measures such as appropriate study design and matching are employed. ***Impact of Regulatory Actions*** Patient and HCP awareness of RMMs - Social media data is well suited for evaluation of risk minimisation measures awareness. Measure Impact over time - Content metadata (time and location) allow investigation of changes in awareness, and thus effectiveness of regulatory actions.	***Supply and availability of medicines*** Shortages - Amount of social media data seems to be linked with medicine shortages with high public interest. ***Public Health Emergency Response Preparation*** Early Detection - The early and rapid detection of health emergencies and associated public health concerns identified from social media can support preparation of regulatory emergency response. ***Misinformation and Stakeholder communication*** Optimising communication - Social media data can be used to map misinformation, which is useful for optimising stakeholder communication and regulatory strategies.

5.1. To support planning of applicant studies

Design and feasibility of planned studies

Data from social media may help the design and feasibility of safety, efficacy and effectiveness studies if such data is leveraged for the recruitment of study participants, when social media are used for data

collection for the characterisation of current and potential users of drugs on the market or in development. In addition, social media has made patient led research more feasible.

Several studies have used social media, both general open platforms as well as patient-focused platforms and disease specific online patient communities to recruit study participants (Coman et al., 2024; Ostler et al., 2024; Ware et al., 2024). These studies have shown that using social media platforms can make screening of eligible participants simple, user friendly and automated. Several studies have also used social media for data collection via surveys and questionnaires to be filled either by the general public or a more targeted population (Altarifi et al., 2024; DasMahapatra et al., 2017; Eindor-Abarbanel et al., 2024). In addition to often being cited as a cost-effective recruitment or data collection method, social media also allow for larger reach and may be particularly useful for reaching patients with rare diseases (Walsh et al., 2022). In fact, Zunic et al. (2020) found that online communities of patients with rare or chronic diseases are among the most active, suggesting the particular benefit of social media data to reach as well as gain insights on the patient experience in these patient groups.

Social media has also been used to further characterise a patient population for a drug in development or already on the market and generate insights into patient. Social media data available on X for example include a geographic marker which have been used to understand the geographic variation of users of different medications, including off label use (Al-Hamid et al., 2024). Use of social media data to understand drug utilisation is further discussed in section 5.2.3.

Another way social media data has been used is to capture the experience of patients who otherwise may not be reflected in the clinical trials used to generate evidence for the authorization of a medicine, for example those who live with more than one disease and use more than one medication. One study exploring the use of Modafinil analysed social media posts related to Modafinil across 5 social media platforms using qualitative and Natural Language Processing (NLP) methods. The researchers were able to discern that almost a third of users have co-morbidities (Walsh et al., 2021). These insights are helpful for understanding the patient population of currently available medicines, which is particularly useful in pharmacovigilance activities as seen in section 5.3.1., but the information can also be used to shape the studies of new medicines and investigate the user population of similar medicines or identify unmet needs in current standard of care as seen in section 5.2.2.

What became apparent through the patient led efforts on Long-COVID is that social media can take on a more central role in medical research instead of solely informing it. Social media makes sharing information and connecting with others across the globe easier than ever. During the COVID-19 pandemic, this allowed patients, caregivers, HCPs, and researchers alike to share the latest knowledge on the novel virus both from the patient experience and clinical outcomes perspectives as well as from early clinical studies and expert knowledge on the field (Callard & Perego, 2021). Social media allowed for patients to not only provide valuable insights into the various symptomologies and prognosis of COVID-19, but to also spear-head and drive the creation of knowledge on the area. Long-COVID as a term originated from online patient discussions. Connections made online between Long-COVID patients and health professionals facilitated planning, funding, and conducting studies on Long-COVID, as well as advocacy to health authorities ultimately leading to Long-COVID becoming a WHO recognised condition (Callard & Perego, 2021). Social media data can thus be used to identify and map ongoing research efforts, but also to facilitate health research and support patient engagement particularly when little research on a given topic exists.

5.2. To understand the clinical context

Support the understanding of disease epidemiology

While social media platforms do not collect clinical measurements per se, social media data can be a complementary source of information to gain important insights into varying disease presentations and prognoses, and into understanding seasonal and geographic disease prevalence.

A review focusing on influenza detection and prediction describes several models that have been developed for epidemiological research using social media data (Alessa & Faezipour, 2018). These models used different approaches from machine learning techniques to text mining and were able to perform spatial and/or temporal monitoring of influenza spread. Similarly, social media data was also used during the COVID-19 pandemic to measure and analyse mentions of COVID-19 or associated symptoms which were found to be in line with the daily reported cases and local outbreaks prediction (Lamsal et al., 2022).

Health, medicine or disease related discussions on social media platforms often occur around patient symptoms. Analysing these patient discussions can give useful insights into the various ways in which diseases are experienced by patients and contribute to a more detailed understanding of symptom clustering. Data from an online health forum was used to understand the clustering of various breast cancer symptoms (Marshall et al., 2016). More recently, data from social media groups of patients with long covid was used to investigate the type of symptom constellations patients experienced with this novel disease (Ziauddeen et al., 2022).

Clinical Management

How patients are treated and cared for (e.g. access to available therapies) can be further understood using social media data.

While electronic health records and pharmacy prescription data describes more accurately the number of times medicines have been prescribed to patients, they often do not capture the reasons why a medicine may not have been prescribed to some patients even though they may have benefitted from them (Y. Li et al., 2011). Patients often turn to social media platforms to discuss experiences with their HCPs and with receiving treatment. A study on Psoriatic Arthritis (PA) patients' experiences analysing reports from various online platforms and including popular social media sites like Facebook and X, described the types of barriers to treatment access, such as high costs and ineligibility for insurance coverage that PA patients face (Sunkureddi et al., 2018). Insights on access to care gathered from social media data can help paint a more complete picture of patient clinical care.

Reynolds et al. (2018) also demonstrated how patients often seek opinions from peers in online patient discussion groups for various reasons, and in doing so describe the advice and treatment they have received from an HCP. Their study showed how patients on patient-focused online groups and platforms describe the clinical care received to a high level of detail, from the prescribed drug dose and regimen to patient medical characteristics and lab test results (Reynolds et al., 2018). While posts on an online discussion do not depict an accurate, exhaustive or complete image of clinical care, they can complement the information collected via other established means (i.e. EHRs, claims data) by providing some patient's perspective related to specific medicines or health conditions.

Drug Utilisation

An area where social media data seems to have particular benefit is in supporting the understanding of real-world medicine use. Social media data can shed light on aspects of patient drug use such as off-label use, alternate drug use patterns and use of multiple substances, that may not be captured during clinic visits or in other RWD sources.

Social media data analysis allows for the examination of a much wider array of drug use patterns than those tested in clinical trials, from different doses to taking medicines on an ad-hoc basis or taking

breaks to increase efficacy (Walsh et al., 2021). A study on patient data from X and Reddit led to identifying novel administration routes for benzodiazepines employed by users as well as poly-drug use trends (Mullin et al., 2024). By analysing user posts on various social media sites, Walsh et al. (2021) found that Modafinil was being used by patients to treat 33 different conditions globally while it has only gained regulatory approval for 4 indications in Europe, with approved use in most other non-EU countries being equally or more limited than in Europe (European Medicines Agency, n.d.). They were also able to retrieve measures of effectiveness, safety, impact on quality of life (QoL), and the various dosages and patterns which patients followed. In fact, they observed that some patients have created usage patterns based on their personal lifestyle and even that smaller doses than indicated in package labels are sufficiently effective for some.

Other studies describe how there is a particular utility for leveraging social media data to better understand abuse and misuse of medicinal products, as nonmedical users of medicinal products are unlikely to report their off-label use to HCPs yet many self-reports can be observed in online discussions (Jain et al., 2020; Nasralah et al., 2020; Sarker et al., 2020). In this manner, social media data has been used to discover themes and topics surrounding drug abuse, understanding public sentiment on- and risk factors for misuse, and factors influencing the spread of online content on off-label use (Jain et al., 2020; Sarker et al., 2020; Young et al., 2020).

Though these findings are compelling, no action by regulatory authorities based on these results were identified. In general, regulatory actions such as labelling changes or calls to withdraw product from the market based on the analysis of social media data were not identified in the reviewed literature, nor in an additional search on the topic. However, this calls for more use of social media and use such findings to trigger a more detailed investigation via other pathways, which could lead to further regulatory action or changes in product label for example.

Consequently, social media data can help regulators to identify unmet needs in current treatments by capturing patient experience data, prescription data and usage patterns, particularly for off-label use. These can be nuanced through the supplementary information on characterisation of patient groups that social media data allows for as described in section 5.1, to understand who or which patient groups are affected.

In addition, social media can be used to investigate why some patients do not start or discontinue treatment, understanding the underlying reasons and make hypothesis about the causes of the unmet needs (Roborel de Climens et al., 2020).

5.3. To investigate associations and impact

Effectiveness and safety studies

Another use-case where social media seems to have a particular utility is in medicines' safety and effectiveness research. A recent report on patient experience data for medicines regulation identified the promising opportunities for social media data to support signal detection and identification of adverse drug reactions (ADRs) in addition to assessing patient risk perception and drug misuse (Almeida et al., 2024). Arguably, the most mature domain in which social media data could readily be leveraged by the European medicines regulatory network is for the detection of adverse drug reactions (ADRs), while it could also offer complementary insights into effectiveness.

Though current ADR reporting systems allow for reliable detection, it is estimated that up to 90% of ADRs go undetected (Dirkson et al., 2022). Particularly less severe ADRs or those which patients may confuse with other common symptoms rarely get reported to HCPs, even though they may have a sizeable impact on the patients QoL. Many patients will look to online forums and social media to seek

information on symptoms they experience, to determine whether they need to go see a professional or to receive advice from other users. As such social media has been recognized as a particularly useful source for ADR detection. Song et al. (2022) retrieved ADRs from WebMD medication user posts finding that more potential ADRs can be found by combining online reviews with ADRs reported through routine pharmacovigilance systems. Another study reviewed and highlighted the cost effectiveness of gathering large amounts of data from social media, and the likelihood for such data to include mild and moderately severe ADRs compared to conventional ADR reporting systems (Walsh et al., 2022). The same review identified social media to be a particularly useful source for identifying unknown drug-to-drug interactions as well as substantiating the knowledge base on drug interactions. Social media can also lead to more rapid detection of ADRs once a product is on the market. A scoping review demonstrated that several studies worldwide have shown social media data to contain novel ADRs from 3 months and up to 9 years before a labelling change enforced by a regulatory agency (Lee et al., 2021). It should be noted that health specific patient forums were more likely to capture ADRs early in general, while for some drugs traditional pharmacovigilance systems were faster. Moreover Gao et al. (2022) suggest that ADR reports from social media accelerate the product recall time meaning that increased number of ADR reports on social media may lead to faster withdrawal of medicinal product from the market. It seems that social media data is useful for early detection of potentially unsafe products for patients.

Social media can also be used to investigate effectiveness of medicinal products to some extent. In the absence of verified clinical tests, measurements of effectiveness from social media data are based on patient self-reports. On some health specific sites like Patients Like Me, users may disclose their laboratory results and clinical diagnoses and these data may be collected through standardized surveys which makes studying effectiveness of a single drug and cross-drug comparisons more feasible (Blaser et al., 2017). More common in spontaneously generated patient experience data (SGOPED) is the presence of more general user posts on the perceived benefit of the medicine by the patient, for example. Several studies noted that causality is difficult to establish based on unverified data or a placebo effect making accurate conclusions on medicine effectiveness nearly impossible to establish (Walsh et al., 2021, 2022; Wessel & Pogrebnyakov, 2024). However, Walsh et al. (2021), suggested that searching for language indicating particularly rapid onset of effects such as "within an hour I started to feel better" can give an understanding of the perceived effectiveness of the medicine. Saha et al. (2019) explored how the use of psychiatric medications can impact mental health through investigating changes in mental health symptoms of social media users as well as their affect and cognition before and after taking psychiatric medication. By simulating a randomized controlled trial (RCT) and adopting a causal inference framework based on matching, they were able to compare the effectiveness of different drugs and drug families as well as analyse effectiveness on individuals.

Impact of Regulatory Actions

The utility of investigating the impact of regulatory actions via social media has been recognised (Brosch et al., 2019). Analysing patient conversations can give an understanding of their knowledge of proper safety measures as well as their sentiment. While the utility of non-individual specific data is limited in most other use-cases, it can imply the scale of the issue by indicating how many people agree or relate. As social media captures a large variety of users across many regions, it seems to have great potential to study the permeability and patient awareness of risk minimisation measures (RMMs) for medicines. Thanks to temporal information embedded to some social media posts, retrospective temporal analysis can be conducted. A study looked at online discussions on X over a period of two years on four therapies for COVID-19, of which two had gained regulatory approval (Hua et al., 2022). The time-stamped data allowed for time-trend analysis and measured the impact of significant events such as regulatory approval or celebrity support of certain medications on the public

perception of available medications. Moreover, as groups and pages targeted for HCPs exist in social media, such analysis can be done to understand awareness of HCPs on appropriate safety measures.

5.4. To monitor, prepare for and address public health challenges

Given the unique features of social media data (further described in section 6.), their utility in regulatory decision making expands beyond the previously identified use-cases for RWD by portraying potential as a valuable tool to support crisis preparedness and identify information of high interest to public health. (European Medicines Agency, 2023).

Supply and availability of medicines

One such use-case is the investigation of availability of medicines, particularly for identifying emerging and ongoing medicines shortages. Anwar et al. (2020) describe how a peak in the amount of social media mentions correlated with a shortage of hydroxychloroquine and chloroquine during the COVID-19 pandemic as it was dubbed online as a 'game-changer' COVID-19 treatment. The shortage impacted the people using these medicines to treat their auto-immune diseases, and led to adverse reactions and deaths in the populations making off-label use of them to treat COVID-19 (Anwar et al., 2020). As another example Cano-Marin et al. (2023) leveraged data from X to explore challenges related to the COVID-19 vaccine supply chain and were able to gain insights into vaccine distribution and demand forecasting for example based on online user generated content. Moreover, it has been identified that the social media mentioning of rapid weight loss properties of the glucagon-like peptide 1 receptor agonist (GLP-1 RA) could be a contributing factor for the global shortage of these drugs. (Han et al., 2024; Pacific, 2023). In instances where the need or demand for a particular product is unusually heightened, as with GLP-1 RA products or COVID-19 Vaccines, social media data could be linked link with the availability of medicinal products (Cano-Marin et al., 2023; Han et al., 2024). To assess the potential use to monitor and potentially predict drug shortages more research applying big data analytics and automated social media mining across various platforms is needed.

Public health emergency response preparation

Early preparation to respond to public health threats is key in their effective and timely management. The need for a well-prepared health care system and thorough emergency response protocols was undeniably demonstrated at the onslaught of the COVID-19 pandemic. Identifying the threat is a key step before launching responsive measures. In the EU, the European centre for disease prevention and control (ECDC) leverages some social media data during routine epidemic intelligence (ECDC, 2024). They monitor data from several sources, including Facebook and X as popular social media platforms, to rapidly detect and assess health events that are related to communicable diseases in the EU/European Economic Area (EEA) member states (ECDC, n.d.). In a similar manner, monitoring social media data can be useful for the EU medicines regulatory network in preparation for emergency responses. In particular, several retrospective studies have shown how signs of infection with the COVID-19 virus could be identified on social media weeks before health care authorities in Europe as well as globally declared the presence of the virus in their respective regions (L. Li et al., 2021; Lopreite et al., 2021; Wang et al., 2020). Moreover, social media data can also give rapid and near real-time insights into associated challenges, for example on monitoring the increase of depression during the pandemic (Zhang et al., 2021). In these ways social media data can be valuable for work on crisis preparedness and epidemic intelligence done in the EU regulatory network.

Misinformation and stakeholder communication

As social media platforms can be accessed by virtually anyone, they act as a common space connecting all stakeholders. Importantly, social media as a communication forum creates a unique opportunity for any user (from patient to industry and regulators) to engage together and to a certain extent influence each other. Understanding relevant online discussions is key for better stakeholder communication. Moreover, since social media data can be generated by anybody, even non-human bots, the prevalence of misinformation can be high and it can be useful to understand how this can influence patients, on their sentiment and awareness regarding medicines and the evidence around their safety and efficacy. Several studies have used NLP and other AI approaches to detect and classify misinformation on medicinal products on social media channels (Edinger et al., 2023; Hossain et al., 2020; Joshi et al., 2023). Mapping misinformation regarding topics relevant to medicine regulation allows for more effective stakeholder communication to combat misinformation and to ensure continuity of public trust on medicinal products.

6. Challenges and opportunities to use social media data in regulatory decision-making

The most commonly reported characteristics of social media data in the literature reviewed for this report are outlined below.

- Social media platforms are hugely popular, with X hosting 206 million daily active users while 52 million daily interactions occur on Reddit (Mullin et al., 2024). In the EU alone X is used by an average of 111.4 million users each month while Reddit hosts 12.8 million active monthly users (*AMARS in the EU*, n.d.; *Digital Services Act (DSA)*, 2024). Social media data accumulates rapidly, and related datasets are large. However, many interactions like reposts of other users' content may not add new information and instead clutter the data set. Many user posts are also short and do not have extensive patient information or contextual details, which poses challenges for the utility of the data for some use-cases.
- The type of content in social media varies largely by platform. While some are built around video or image sharing such as Youtube or TikTok, others feature mostly text, e.g. reddit or X (formerly known as Twitter) (Walsh et al., 2022). Some like Instagram allow users to browse new content and search for any other registered user, while others, like Whatsapp or Telegram, create a more closed ecosystem. Each platform embodies a mixture of the aforementioned features to different extents, and direct user to user communication is common across platforms. Moreover, different social media platforms are favoured by and thus cater to different userbases which impacts the type of content available on any given platform. TikTok and Snapchat are more popular with younger audiences for example, while Facebook and LinkedIn are also popular among older populations. Social media as a (real world) data source is described in more detail in Annex 8.4.
- Social media data encompasses all data generated via social media, and the proportion of irrelevant data, including false data, is high. Data relevant to regulatory decision-making could be a part of any social media post, though most of the relevant data for regulatory decision-making is generated in patient-focused forums, in online groups of patients with the same condition or health specific social media platform such as PatientsLikeMe. The relevance and fitness of data will depend on the regulatory use-case and the research question at hand.
- Social media data is present in various formats depending on the type of user interaction and the type of social media platform. Social media data can be related to user profile information (e.g. username, age, and gender), original user generated content (e.g. distinct video, text, or image posts), online user activity (e.g. time spent on the platform, direct user messages, and connections

to other platform users/pages), groups and communities (e.g. public or private user groups and their members and moderators), content metadata (e.g. geotag and timestamp), and content engagement metrics (e.g. likes, shares, and views) for example.

- Free-text data is common across platforms and can be posted alone or to accompany visual content (e.g. videos and images). Image, video and text data are mostly unstructured and thus will bring specific challenges. But most social media platforms also collect structured data such as the number of "views" indicating the popularity or spread of that post. Another example are simple engagements with user content through "likes" or "shares" indicating how other users react to it. Additionally, some platforms collect user data in a structured way via standardized questionnaires or polls.
- Social media data captures various patient information from demographic and geographic information to patient experience data. Patients often share impacts of treatments or diseases on their QoL, specifically how their mood and emotions, social relationships, employment and daily functioning has changed (Walsh et al., 2021). To varying degrees, social media data can also give indications of the medical history of a patient, i.e. how long they have been using a drug, experiencing symptoms or co-morbidities.
- Social media is widely used globally across different populations, however difference in use patterns exist. Wealthier and younger people are more likely to use social media, impacting the representativeness of the available data. Some studies have found women to be overrepresented specifically in posting health related content (Anderson et al., 2024; Park et al., 2022). However, these studies relied on content from social media posts instead of user profile information and with many users not explicitly stating their gender, it is not possible to ascertain the actual breakdown of users' genders.
- A unique feature of social media is its potential to capture data from populations who have not yet been classified as patients in the health care system, for example by identifying and predicting influenza infections early on (Athanasiou et al., 2023; Velardi et al., 2014).
- Another particular feature of social media as a RWD source is the relative speed at which a data entry can be created and thus how early it can reach regulatory authorities. The administrative actions preceding a social media post in most cases is limited to user log-in while updating an EHR or patient registry may require additional steps such as data processing, sharing, and logging as well as human resources and thus can exhibit a more significant time-lag. Data generation on social media is less restricted by temporal and logistical restrictions as institutionalized RWD sources.
- Access to social media data is increasingly restricted and depends on the platform and the type of data sought after. Data posted in private groups or platforms protected by user login may be inaccessible or more difficult to make use of in line with the general data protection regulation (GDPR). X and Facebook used to be some of the main sources of social media data, however due to recent changes in Facebook and X's privacy policies, less data is available and freely accessible for research.

6.1. Operational challenges and opportunities

Challenges of operational nature include barriers to access to social media data, ethical use of such data and patient data protection, as well as limited guidance due to little to no prior use of social media data in regulatory decision making.

Though social media platforms host vast amounts of data, access to social media data is unstable and restricted. First of all, activity taking place on social media and the operation of the platforms themselves are situated in a greater socio-political and economic context creating a volatile landscape for data availability and access. For example, TikTok, the popular social media app has faced scrutiny over privacy concerns leading several governments issuing bans for the use and operation of the platform (Fjallhed et al., 2023). As another example, previously open and public social media groups can become closed private groups limiting both access to the data as well as its usability under the GDPR as non-public data (Walsh et al., 2021). This in turn limits the analysis and replicability of some studies using social media data. Moreover, as companies continue to capitalize on the profitability of patient data, free and open access for health research has become a rarity in the market for social media data. Platforms like X and Facebook, which have been the go-to sources of social media data for research due to easy access to large data sets via their application programming interfaces (APIs), have now blocked free access or drastically reduced the number of users' posts available (Walsh et al., 2022). While these trends are influenced by economic incentives, increased emphasis on data privacy has also called for restricted general access and strengthened data protection measures. Even though access to data from previously useful social media platforms is currently restricted and may change in the future, data that has already been collected can still be useful, particularly in training Artificial intelligence (AI) models to mine other platforms (Klein et al., 2024).

The secondary use of patient data posted online is coated with complex ethical considerations. Social media sites collect identifiable personal data such as username, IP address, and location from users when creating a social media post, even on public sites. Several studies have exhibited ways to either anonymise or pseudonymise these identifiable data or not collect these data in the first place (Al-Hamid et al., 2024; Anderson et al., 2024; Sunkureddi et al., 2018; Walsh et al., 2021). However, as the paper by Wessel and Pogrebnyakov (2024) shows, even data of interest, such as symptoms, patient gender and diagnostic status can be traced to a single user and thus a risk of patient re-identification remains. Re-identification attacks are a documented patient data privacy issue (Emam et al., 2011). More reliable ways to eliminate these data privacy concerns are needed.

Though a convincing case is made for social media recruitment, advertising clinical trials through social media has significant ethical issues and can undermine social media users' privacy (Mühlhoff & Willem, 2023). Advanced algorithms and machine learning techniques are currently leveraged on many social media platforms for targeted advertising, to identify potentially eligible participants for clinical trials (Mühlhoff & Willem, 2023). In doing so, large amounts of a user's behavioural data is being collected, sold, and used to make predictions around sensitive health related information like disease status or behaviours risk factors to determine the user's likelihood of being an eligible participant (Mühlhoff & Willem, 2023). Not only does this violate the user's privacy by making assumptions and disclosing sensitive health information to the platform, but it also puts other users at risk as these data are used to train models which could be applied to other users for different conditions and purposes. Moreover, targeted advertising may impact research quality by mending the sample selection. While these issues are significant, they seem to be specific to targeted advertising making use of behavioural data and predictive algorithms. Moreover, some platforms such as Facebook have banned health related advertising to an extent which also restricts advertising possibilities for clinical trials (*Facebook Ad Ban May Squelch Medical Research Recruitment*, 2022). Further ethical and methodological consideration is needed, however, seemingly social media data could still, to a limited capacity, be helpful via non-targeted advertising methods on some platforms.

Furthermore, the existing regulations and terms and conditions of social media platforms must be followed. Chiauzzi and Wicks (2019), describe how unauthorized scraping of social media data and referencing of users in closed groups without appropriate consent or ethical approval have occurred in

social media research. Users expect different levels of privacy across different platforms and platform functionalities. Users expect more privacy, in terms of how many and which other users are able to view a given users data and posts, on password protected social media sites compared to those where user registration is not required (Walsh et al., 2021). In similar fashion, users expect more privacy and restricted viewership for the content they may post in closed groups compared to public forums. This begs for critical evaluation of patient consent and the ethics around using social media data, particularly from closed groups, each time the use of social media data is considered to address a (regulatory) research question. Still, opportunities exist in leveraging platforms geared for the public discussion of health topics or those which are run or supported by patient organizations or are otherwise trusted entities aiming to bring together patient voices to be heard by the healthcare system (Walsh et al., 2021).

6.2. Technical challenges and opportunities

Technical challenges related to social media data include data completeness, volume, reliability, verifiability and processing which inform and influence its quality.

Patient data collected from social media often suffers from incompleteness. On social media platforms, patients may not describe the medication to a sufficiently granular level (i.e. dose and brand name) and they may leave out other relevant factors such as demographic information, personal risk factors, and other medical history. Social media data may not always provide the complete context of medication use, nor be specific enough to address some research questions. The data may turn out to be unsuitable for being regarded as a stand-alone source for RWD or to replace any of the existing RWD sources. As showed by this report, insights from social media data can however complement evidence generated through other methods in areas of post-authorisation safety monitoring, and it can be used to describe patient experiences and contribute to inform decision-making in a variety of domains.

But it should be noted that the necessity for completeness decreases when social media data is utilised as an initial approach either to flag novel ADRs or to support the planning of clinical trials and non-interventional studies for example, ultimately functioning as a springboard for further research. Dirkson et al. (2022) highlighted how some of the key advantages of social media are the unparalleled volume of data, near real-time monitoring, and leaning into an informal environment where patients are more likely to share information with peers than with HCPs. Moreover, inspecting patients' social media posts allows for direct observation of patient perspectives (Walsh et al., 2021). Any patient perspectives observed from other RWD sources such as EHRs have been filtered through an HCP or another party inputting and sharing the data. In addition, social media has the advantage of capturing ADRs not reported through other systems, where trivial or common ADRs for example may go underreported due to limited time of HCPs (Dirkson et al., 2022). In addition, all medicinal products do not enjoy the same amount of attention and discourse on social media, opening room for different types of biases in the use and interpretation of social media data. Products for psychiatric indications have a high number of mentions across platforms in general (van Stekelenborg et al., 2019). Differences can be observed between platforms and type of data as well. Higher volumes of posts discussing ADRs for cancer treatments and medicines with an orphan designation are more likely to be found on health specific patient forums and platforms compared to general sites like X (van Stekelenborg et al., 2019). In addition, social media trends as well as drug marketing can influence the volume of data for a particular medicine. A glucagon-like peptide 1 receptor agonists (GLP-1 RA) approved for the treatment of diabetes became viral on social media, with posts being viewed by millions, and was promoted by several social media users for its weight loss effects (Basch et al., 2023). As such, the analysis of social media data for some drugs is more beneficial to some than

others, due to the apparent differences in availability of data, and it requires careful contextualisation and appropriate methodologies to account for potential biases and confounding factors.

The rapid generation of high volumes of data on social media creates challenges around data processing. The large data sets require more advanced and highly performing analytical infrastructure. The recent years have seen major developments in computational tools available to deal with Big Data, and thus currently available methods, particularly AI powered approaches render the large volumes as an opportunity to be harvested instead of a limiting challenge.

Social media data is also often crowded with irrelevant content, duplicates in the form of re-posts, or fabricated content which may skew the available data. Social bots, automated social media accounts run by computer algorithms as opposed to humans, have become increasingly sophisticated and more and more common across social media platforms (Gorodnichenko et al., 2021). On X, the platform on which most literature on social bots has focused on, it is estimated that around 9-15% of all accounts are automated, which equals to almost 50 million accounts (Hayawi et al., 2023). Social bots can be benign, and automatically post-earthquake alerts when one is detected, they can be neutral and post jokes or nonsense, or they can be malicious and post fake news or spam (Orabi et al., 2020). Regardless of the type of content social bots post, all their posts are by non-human users and thus these accounts must be identified so that their contribution can be restricted in a social media data sample. Several deep learning and other AI methods have been developed for the identification of social bots, though these methods ought to be continually refined with social bots becoming more sophisticated (Hayawi et al., 2023; Orabi et al., 2020).

The informal use of language leads to colloquialisms and complexities with semantics as well as alternate or misspellings of words (Walsh et al., 2022). However, various data processing and analytical tools have been widely researched and many can effectively deal with unstructured, large, and incomplete datasets. NLP tools are particularly useful for the appropriate translation of social media jargon to MeDRA terms. While the accuracy of these tools seems high, their utility for regulatory uses beyond ADR identification need to be assessed.

In addition, data types other than text data from social media have been less researched for patient data. With the popularity of video based social media platforms like TikTok and Youtube, more research is needed to leverage their potential for regulatory decision making. Additionally, methods to overcome the particular challenges such as processing of video, image and audio data, low image resolution, differences in accents and tones, to name a few, need to be developed. As more AI powered tools will be required to process visual and audio data, challenges related to bias in developing AI tools must also be accounted for.

Lastly, the inability to verify the reliability of information described by patients as well as their motives is a challenge which limits the utility of social media data. Particularly when data is anonymised there is no appropriate way of contacting the user or cross checking with other sources. Social media users are mostly not HCPs and thus their descriptions may not always be accurate. In addition, more and more pharmaceutical companies are investing in direct-to-consumer marketing on social media platforms by paying social media ‘influencers’, users with a substantial followers base, to promote their product (Willis & Delbaere, 2022). Recent years have seen an increase in the amount of drug related social media posts where the underlying financial incentive is not always appropriately disclosed (Willis & Delbaere, 2022). Social media data that has been generated because of a paid partnership should not be used or considered as patient data. The lack of transparency likely makes it more difficult to discern between paid partnerships and spontaneous user posts, and thus more research and development of methods capable of identifying promotional material where it may not be obvious is needed. Fake accounts, or bots, may also be creating content on social media leading to a skewed and

inaccurate view of the actual situation. However, it has been shown that many users on health forums match real patients and their medical records (Eichler et al., 2016). Combining data from already established sources with the novel insights from social media could allow to contextualise and verify reported information or control for biases for example. Additional measures can be taken to minimize the impact of some false posts by analysing large data sets and prioritising health related social media sites over general mainstream sites which promote posts that gain a lot of reactions instead of accuracy of information.

6.3. Methodological challenges and opportunities

In terms of methodological challenges, lack of precedent as RWD source as well as the representativity, bias and lack of verifiability of social media data curbs its utility across regulatory use-cases.

As prior use of social media data in medicine regulation is limited, preceding experience to draw learnings from as well as guidance around leveraging social media data are insufficient. For the time being, existing support mechanisms, namely the EMA's Innovation Task Force (ITF) meetings and scientific advice (SA) should be leveraged by applicants planning to utilise social media data.

Social media data, though vast, does not accurately represent the general population. Users tend to be younger and wealthier overall and women and more highly educated people are more likely to make use of patient-focused social media communities (Dirkson et al., 2022). Though accessibility features exist, people who have too little functional capability for example may be unable to use social media, thus entire patient groups may not be represented in social media data. Moreover, users who feel more strongly about a product may be more vocal than other users, leading to a skewed representation of the patient population. Measures to ensure data privacy, such as anonymising data or not collecting identifiable data in the first place, may also narrow the researcher's ability to determine the representativeness of the sample, when factors like location and gender may be lost.

On the contrary, social media data may in fact be well suited to gather information about particular patient groups otherwise underrepresented in health research. A pertinent case is to better understand drug use behaviour and the patient experience of individuals who are pregnant or lactating. As insufficient data exists on the safety and effectiveness of many medicines for pregnant people due to limited inclusion in clinical trials, the importance of data from other sources is heightened (Gelder et al., 2019). Rezaallah et al. (2019) used social media data to explore experiences and perceptions of patients with multiple sclerosis regarding medication use during pregnancy. They were able to identify topics related to disease progression, drug related concerns, advice giving and recounting of health care experiences for example. Moreover, Gelder et al. (2019) used social media data to understand pregnancy related patient perceptions through comparatively investigating differences in perceptions of safety of medicines in pregnant women and local safety classifications. They found some vast differences in perceived safety and safety classifications according to the classification scheme. There seem to be particular opportunities for leveraging social media data to further understand relevant aspects of patient experiences of underrepresented patient groups, such as pregnant and lactating individuals.

Representativity bias also remains an issue for other modes of evidence generation from clinical trials to EHRs. In fact, one of the relative strengths of social media data is that patients who otherwise might be included in clinical trials, i.e. patients with co-morbidities or who use several medicines, are able to make their contribution through social media data. Thus, opportunities exist to benefit from the unique array of patients on social media to complement data from other sources.

Selection biases also impact the utility of social media data as some patients may not be comfortable discussing personal topics related to health in public settings such as open social media platforms for example. Thus, social media data should not be exclusively relied on. However, in general, users of social media and online discussion forums are willing to share personal medical information, particularly when it can be done anonymously (Roborel de Climens et al., 2020). This could be due to a variety of reasons, for example given the expectation that they will get personalised help from the community, or because they are more comfortable sharing certain aspects of their behaviour and drug use with perceived peers than with authoritative figures during clinical visits (Dirkson et al., 2022; Puspitasari & Firdauzy, 2019). Moreover, in some cases social media may capture important conversations that would otherwise not be reflected in health system records, for example when a particular disease or condition is stigmatized. To illustrate, sexually transmitted diseases (STDs) and HIV/AIDS remain stigmatized in various communities, and particularly in smaller cities (White Hughto et al., 2017). Social media sites, specifically those which allow for user anonymity, foster a space where patients can engage in discussion around difficult or stigmatized topics without a risk of personal identification and judgement. In this way opportunities exist in leveraging social media data where other RWD sources may be expected to fall for such non-response bias.

Social media data hosts potential unlike many other data sources to get population level data on off-label use and alternative therapies directly from patients and in near real-time. As social media data is not generated in a controlled environment, the lack of posts about the topic of interest, for example an ADR, does not mean that it did not occur. When exploring medication use, most patients only discuss the current situation they face and may not describe medications they may have used but stopped due to side-effects for example (Blaser et al., 2017).

7. Points for consideration for future actions

7.1. Operational points for consideration for future actions

1. Expedite access to social media data

For the regulatory system to adopt more social media data in its decision-making, pathways enabling access to social media data must first be established.

- **Social media data sourced from private companies:** Some social media platforms provide data for research purposes which could be leveraged for insights into PED or other data relevant for regulatory decision-making. Several third-party companies also provide social media mining services which could provide a simple solution for collection and analysis of social media data.
- **Social media data collected by the EMRN:** Alternatively, the regulatory network could consider setting up network social media data collection, potentially leveraging AI tools for social media mining.

Proof-of-concept programs to access social media data and generate evidence from RWD and/or PED through the aforementioned pathways should be established. This might be particularly useful to address wider public health issues such as international drug shortages or preparing responses to epidemics. Examples of studies for EMA scientific committees are given in 7.2.2.

2. Increase discoverability of RWD studies and sources utilising social media data

In the recently launched HMA-EMA catalogues for RWD studies and sources, it is currently not possible for users to filter for RWD sources or studies using social media data. For the time being, registrants of

RWD studies are only able to indicate the use of social media data under the category 'other,' but this is not reflected in the filtering criteria for the end-user interface. For RWD sources, it is currently not possible to register a source as containing 'social media data' nor is it possible to specifically search for sources containing social media data. To increase discoverability, the RWD catalogue for studies should better depict when a study has utilized data collected from social media by considering specific tags or filters for social media data. The catalogue for RWD sources should be updated with features allowing the explicit indication of social media data use both in the registration of a data source as well as in the search filters available in the end-user interface.

3. *Engage and collaborate with all actors and relevant initiatives/research in the healthcare sector and leverage ongoing initiatives on PED*

Much of social media data falls under the umbrella of Patient Experience Data (PED) as social media provides a platform for patients to express their concerns and share their experiences with diseases and treatments with peers, experts and the wider scientific community. Social media platforms can provide crucial and early PED where scientific literature is still emerging, as seen with post-COVID-19 conditions or long COVID. The EU network is in the process of developing a reflection paper on the best EU approaches for defining, generating, collecting and analysing PED. This work should leverage social media data to support the use and establish the value of PED and social media data for regulatory decision-making.

Early and continuous engagement with all stakeholders is essential for building skills and competencies, and laying a solid foundation for the use and adoption of social media data. Patients and patient organisations' expertise and knowledge should be leveraged when considering use of spontaneously generated online patient experience data (SGOPED) to ensure that the patient's messages are interpreted and applied in a meaningful way. Furthermore, some patient organisations manage their own patient-focused platforms or social media groups meaning partnerships can facilitate access to more targeted and reliable patient social media data. Moreover, as many HCPs participate in online discussions and generate social media data while also playing a key role in the clinical care of patients, their engagement is crucial. In addition, coordination with the pharmaceutical industry is necessary to ensure alignment with current industry practices.

Furthermore, despite lacking regulatory precedent in the EU, valuable work on the field of social media research has already been conducted from which learnings can be used to strengthen regulatory processes. For example, the WEBRADR initiatives' both editions have been fundamental in establishing the basis for further use of social media data, while more recently initiatives like the Pistoia Alliances' initiative "Social Media to Produce Real World Evidence (RWE)" have worked to advance the utility of social media data in medicine regulation and drug development (Maskell, 2022). Identification of relevant initiatives is rudimentary to leveraging their contributions. Moreover, to leverage the research skills present in academia, social media data and the areas where this report identified further research needs should be more strongly represented in the EMA's regulatory science research needs and/or at EU level.

Ultimately it is proposed to organise a workshop on social media data to discuss use case, share learnings from initiatives and experts, and agree on a common network position and roadmap to further progress the use adoption of social media data in EU medicine regulation.

4. Draft a guiding principle on ethical considerations for the use of patient data from social media.

Compliance to data protection requirements is expected by default. But as discussed in this report, social media data use comes with complex ethical challenges. Patient consent, use of sensitive patient information and the risk of re-identification warrant detailed assessment of the ethics of the EMRN leveraging social media data in regulatory decision making. These findings are to be translated into guiding principles for the ethical use of social media data in regulatory decision making. Patient involvement is urged in the development of the guiding principles, learnings from proof of concepts studies. To minimize potential data protection issues, data from publicly available online discussions, with a focus on platforms that are freely accessible without a log in barrier, should be prioritised in any proof-of-concept studies.

7.2. Technical points for consideration for future actions

1. Launch and evaluate proof-of-concept studies on social media data for regulatory decision making

As social media data is yet to be widely accepted in EU medicines regulation, proof-of concept studies on the use of such data for the different regulatory use-cases outlined in this report should be considered to help build experience and learnings. Potential topics are listed hereafter.

Given the unparalleled reach of social media, as well as already existing patient groups for a wide array of conditions, the committee for orphan medicinal products (COMP) could investigate the added value of social media data to justify significant benefit to patients. More specifically, monitoring online discussions on rare diseases and relevant treatments can give insights to patient perspectives on the burden of treatments, including potential burden of product administration and unmet needs in current standard of care. These findings may be particularly helpful to substantiate claims for major contribution to patient care and thus be useful in the context of orphan designation assessments.

As young people are overrepresented particularly on generic social media platforms, these platforms may seem to carry potential as an abundant source for paediatric PED. The paediatric committee (PDCO) could investigate the use of social media to further understand children’s sentiments as well as drug use patterns and off-label use for example. Platforms, such as TikTok, Reddit and Instagram, favoured by minors should be prioritised.

Considering the unique potential of leveraging social media data to better understand drug abuse and misuse, the pharmacovigilance risk assessment committee (PRAC) could consider exploratory projects to analyse online discussions around inappropriate use of medicinal products. Such analysis is likely to be beneficial in assessing the success of RMMs. Particular attention should be placed on how social media data may supplement insights from other data sources given its unique characteristics and capacity to amplify the patient experience.

As another example, the regulatory network could pilot the use of social media data to prepare for public health emergencies responses and monitor medicine shortages. Social media data across the EU member states would be valuable in this, and thus cross regulatory collaboration would be particularly beneficial for this. Moreover, given rapid developments in AI methods, the regulatory network should consider launching initial pilots to assess the feasibility of AI powered approaches to mine social media platforms for regulatory quality patient data.

Later on, findings from these proof-of-concept studies and the workshops proposed in section 7.1.2. could be used to launch more established projects. Ultimately, the results could culminate in the development of network principles to guide when and how social media data could be leveraged, and which social media data sources are fit-for-purpose for different regulatory use-cases. As social media sites are many, with new platforms continuously being created and the amount of content on different

medicinal products on social media varies significantly, guidance specific to social media could be considered. Therefore, the workshop and pilot learnings could be used to build on the currently available suitability criteria for RWD data-sources which are laid out in the Good Practice Guide for the use of the Metadata Catalogue of Real-World Data Sources (European Medicines Agency, 2022b).

7.3. Methodological points for consideration for future actions

1. Support projects for understanding the current use and development and validation of new methods for analysing social media data

For social media to be adopted more widely, confidence in the analytical methods must be fostered as well as their use in regulatory procedures up to date must be better understood. While some reporting on the use of social media data has been conducted, particularly in pharmacovigilance activities, a detailed exploration on the use of social media data in the pre-submission phase and at time of benefit risk assessment would be fundamental. This research should investigate how social media data may have been used in previous MAAs as well as a comprehensive exploration on discussions of social media in SA and ITF meetings. For example, whether social media data was used to support participant recruitment, or to refine study design based on PED. Once the current state has appropriately been described, a more detailed understanding of the regulatory position regarding social media data is expected to be achieved. Subsequently, such an evaluation will be useful for validating where additional regulatory efforts should be aimed. Therefore, such review could be included in the next revision of EMA's regulatory science research needs by identifying knowledge gaps and evaluating the reasons for or against use of social media data in MAAs or early interactions so far.

Further research to develop and validate analytical methods to draw insights from social media is to be facilitated. Notably, both academia and industry have made great progress in developing efficient and precise tools powered by AI techniques, as seen in the work presented in the annual Social Media Mining for Health Applications (SMM4H) symposiums (Klein et al., 2024). Regulatory bodies must ensure up to date knowledge of the currently available analytical methods and of their clinical relevance and reliability for regulatory needs. The increased knowledge on AI in the regulatory network through the work done on AI in the BDSG will be crucial to better prepare for the use of social media in more regulatory uses. As such, the BDSG should consider the further exploration of AI approaches and methods leveraging social media data as a case-study or focus.

8. Annexes

8.1. Search Strategy & Characteristics of Search Results

The literature search on Embase yielded 4091 articles and 549 reviews as results coming to a total of 4640. In addition, key word searching was used to find articles related to specific topics emerging from the literature search such as "social media and product recall." Articles were reviewed if their research focus was evaluating social media data or related aspects relevant to its use or collection as RWD, such as papers developing NLP tools to gather social media data. Studies were excluded if their research focus was solely on social media as a health intervention, for example social media as a health promotion intervention.

The most relevant and recent articles were prioritized for this review, due to the large volume of results. Using the "sort by relevance" filter on Embase the first 500 abstracts of the search results were scanned, though after roughly 300 articles, a point of saturation was reached meaning the remaining papers did not contain new use-cases for social media data, or new information on data collection or analysis methods relevant to the scope of this paper. To ensure that appropriate representation was achieved, every 10th page of the remaining search results was checked but no new relevant concepts were discovered. Using the "sort by date" filter, the abstracts of the first 500 of the most recent articles were scanned to ensure the inclusion of most recent developments otherwise not captured by the currently available systematic reviews and meta-analyses. In addition, a free internet search was conducted to look for initiatives relevant to social media data.

8.2. Search Terms

Subject Head Search Terms:
'social media' OR instagram OR 'social platform*' OR whatsapp OR youtube OR 'social network*' OR reddit OR pinterest OR myspace OR linkedin OR tumblr OR twitter OR facebook OR tiktok OR 'online patient*' OR 'online community' OR blog* OR 'e-communit*' OR 'digital patient*' OR 'digital communit*' OR 'e-patient*' OR 'social media monitoring'

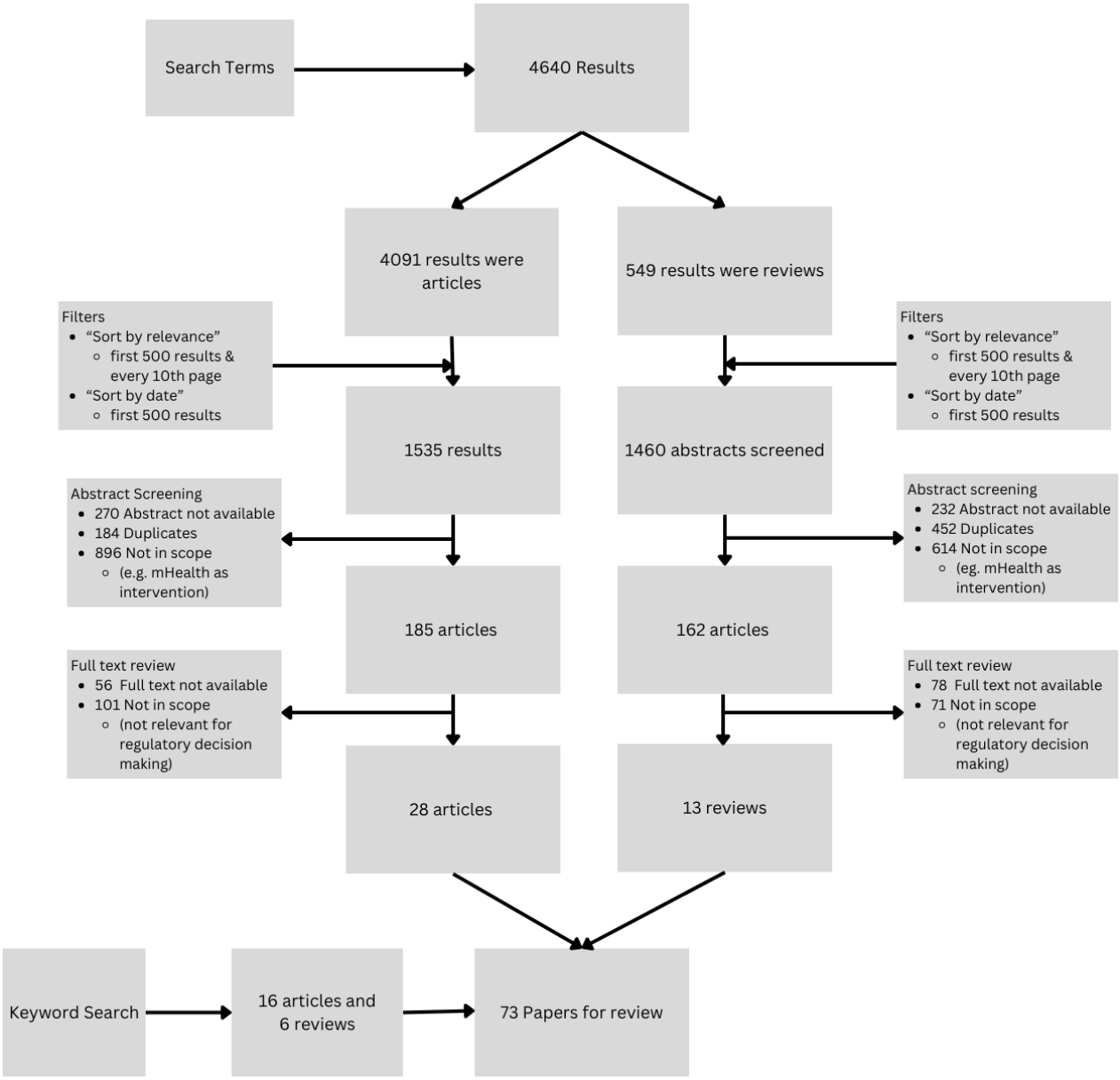
AND

'medicine regulation' OR pharmacovigilance OR 'adverse drug reaction' OR regulator* OR 'regulatory decision making' OR 'clinical decision making' OR 'medical decision making' OR 'shared decision making' OR 'drug* regulation*' OR 'narcotic* control*' OR 'drug* control*' OR 'marketing authori?ation' OR 'drug* approval*' OR 'approval procedure*' OR 'approval process*' OR 'medicine* approval*' OR 'drug* assessment*' OR 'medicine* assessment*' OR 'drug* evaluation*' OR 'medicine* evaluation*' OR 'european medicines agency' OR 'food and drug administration' OR 'adverse drug event*' OR 'side effect*' OR 'adverse reaction*' OR 'adverse event*' OR 'drug monitoring' OR 'drug surveillance' OR 'post marketing authorization' OR 'drug safety' OR 'benefit-risk' OR 'risk assessment' OR 'drug efficacy' OR 'drug effectiveness' OR 'drug use' OR 'epidemiology' OR 'disease course' OR 'off label drug use' OR 'treatment pattern' OR 'drug therapy' OR 'drug dose' OR 'patient reported outcome' OR 'medical record'

Key Word Searches

'Social media data in pharmacovigilance'
'Social media and labelling change or product recall'
'Social media data secondary use health'

8.2.1. Search strategy visualisation



8.3. Tables

OPTIMAL	Challenge	Opportunity
Operational	**Access** Data access is unstable due to: - Political restrictions - Company (privacy) policy changes - Public groups may become private	- Leveraging already collected data to gain insights into the various use-cases despite limited access - Anonymising available social media data / using methods (AI powered) to not collect personal identifiers
	Data Protection & Ethics - Much SM data is sensitive (health) data	- Employing appropriate existing methods to ensure privacy (and stop re-identification) in processing and secondary use of SM data.

OPTIMAL	Challenge	Opportunity
	- Stricter protection measures and ethical considerations - Risk of Re-identification - (Un)availability of patient consent & Expectations of privacy depending on the platform	- Prioritise use-cases that are more clearly ethically aligned with applicable ethical principles, e.g. leveraging public discussions and/or those facilitated by patient organizations
	Limited Guidance - Lack of prior experience - Grey areas on ADR reporting responsibility	- More specific guidance, SM data quality reflection paper - Continue and implement WEBRADR and other related initiative's work
Technical	**Incompleteness** - User posts not always detailed enough - Patients may not describe past treatment and medical history	- Making use of SM as a supplementary data source - Using large datasets to mitigate missing data
	Varied amount of data - Some medicinal products are discussed less in social media leading to less data	- Using social media for those medicinal products which have a lot of data
	Reliability - Presence of irrelevant data - Duplicate posts (reposts) - Bot posts - Limited scope to verify reliability of data ▪ Data used should be anonymised which makes follow up difficult - Some date is based on financial incentives	- Prioritising patient focused platforms as many users there match real patients - Leveraging advanced (AI) tools, e.g. to process/clean data and better identify and exclude irrelevant and promotional content
	Quality - Misspellings and colloquialism - Language, Current AI tools lack behind in other languages than English (though rapidly evolving) - Video, image and audio data analysis less researched than text data	- Opportunity to leverage latest AI tools, pretrained models which get retrained on the specific data set, which can effectively fix misspellings and understand abbreviations.
Methodological	**Limited Regulatory Experience and Precedent** - Little to no prior EU regulatory experience may hinder regulators confidence in social media data	- Leveraging existing support mechanisms like ITF meetings and SA to address concerns
	Population Representativeness - Age and social status of users of social media platforms can vary greatly and introduce biases to consider when analysing data, e.g. older and less wealthy patients are underrepresented in general - Women are overrepresented on patient focused platforms	- Accessing more & different patients than CTs - Leveraging social media for areas where it holds unique potential (e.g. PED, stigmatized topic like STDs, or misinformation), not for general conclusions about safety and efficacy

OPTIMAL	Challenge	Opportunity
	Patients may not discuss all topics in public settings	

Points for consideration for future actions

OPTIMAL	Points for consideration for future actions	Core Action	Strategic Goal
Operational	Expedite access to social media data	Facilitate access to social media data by exploring the two main pathways for data access: - Through social media companies or external contractors - Through an internal work program and/or team focused on social media patient data collection and analysis	To pave way for the efficient incorporation of social media data into the routine regulatory decision-making processes for which social media data can be beneficial.
	Increase discoverability of RWD studies and sources utilising social media data	Options to tag RWD sources and studies that utilise social media data should be included in the RWD catalogues.	Making the RWD studies and sources that include social media data more explicit, transparency and discoverability is increased.
	Engage and collaborate with all actors in the healthcare sector and leverage ongoing initiatives on PED	Promote early interaction with regulators. Leverage social media in PED work and explore use of social media where scientific literature is still emerging. Organise a workshop to involve patients in the development of a road map and agree on the EMRN vision for the use of social media data in regulatory decision making.	Patient involvement ensures correct interpretation and meaningful application of social media data in regulatory decisions. Alignment with industry is crucial for effective use of social media data in the EMRN and promoting and engaging with existing initiatives supports a stronger network.
	Draft a guiding principle on the ethical considerations of use of patient data from social media	Evaluation of the ethics of EMRN use of social media data should be pursued. Ultimately an ethical guiding principle should be developed, outlining the ethical considerations to be considered when making use of social media data.	Establishing a strong ethically sound foundation for the use of social media data is essential to ensure patient and EMRN trust given the large amount of sensitive data, and thus risk to data privacy that accompanies its use.
Technical	Assess the need for specific data quality guidance for social media data	Consider whether social media data should be included in the DQF or as an additional chapter or guidance, with a particular focus on pharmacovigilance.	Given the particular potential of social media for enhanced safety monitoring, additional efforts to ensure high quality social media data may increase the regulatory acceptability of social media data.
	Launch and evaluate pilots on social media data for regulatory decision making	Pilot projects for the various proposed use-cases should be conducted. For example, for the investigation of different presentations of rare disease or for paediatric indications.	Pilots on social media data use will build the necessary foundation for the further adoption of social media data into routine regulatory decision making.

OPTIMAL	Points for consideration for future actions	Core Action	Strategic Goal
Methodological	Support initiatives for further development and validation of new methods to analyse social media data	In depth assessments of previous MAAs, SA and ITF meetings should be conducted. Support network initiatives to investigate and develop new methods for analysing social media data, e.g. considering analytical biases around age, sex, social status.	These studies will give valuable input on the capacity to which social media has been and could be implemented into existing regulatory processes.

8.4. Social Media as a RWD Source

Various kinds of social media platforms exist with an array of different features and thus different available RWD. There are so-called general social networking sites or general social media platforms where usership or user interactions are not limited to a specific topic. There are also platforms created specifically for health-related topics, or patient focused platforms, for patients with a particular disease for example. For social media platforms to offer utility for regulatory decision making, the data must be relevant and accessible among other necessary data quality aspects. Given the diversity of social media platforms in use by patients globally it is necessary to understand the features which shape the utility of data available on social media. Moreover, the private or public status of the data is a key consideration for its secondary use.

8.4.1. Common features of social media platforms

The type of available features and user content varies between social media platforms. However some common features that are found in most platforms are distinguished, but not limited to the user profile, network, messaging and content stream (Bayer et al., 2020).

The user profile represents the individual using the social media platform. Here the user can update personal information such as age, name, location, and personal status updates for example. As such the user profile can be used to gather demographic data of patients as well as data relating to the individual's medical history and drug use if they have posted relevant status updates. Depending on the platform and user preferences, however, some user profiles can be publicly accessible while some can be private, only accessible by the user's contacts.

The network represents the web of connections the user engages with and includes a user's contact lists and 'tags' or 'mentions' of or by other users. Depending on the platform user connections can be one-way 'following' or two-way 'friends' or 'contacts'. Analysing the network, data on the social behaviour and connectivity of users can be analysed, which could give insights on disease spread for example.

Messaging represents the direct communication between user profiles. Some popular platforms such as SnapChat, Whatsapp, and Telegram centre this feature while other platforms such as Instagram or Facebook include it as supplementary feature. Since direct messages are meant only for the recipient(s), some patients may be more likely to share sensitive patient information. However, as these data are not public, access is highly restricted and their mining as RWD would be a breach of ethics and data privacy.

The stream refers to the flow of content on social media platforms that is made up of posts from different accounts, as well as the reactions and comments to these posts. The stream a user observes is curated to that user as it consists of content by the accounts in the user's network, as well as sponsored or otherwise promoted content by the platform's algorithms. Though the exact selection and order of posts is specific to an individual user, the content reflected on the stream draws on user content available in a pool of content on the platform's servers. These user posts can also often be accessed by any user through appropriate key words and tags as long as they have access rights. From a health care data perspective, the stream features public user posts which can be mined or analysed for the collection of social media patient data to understand unmet needs or patient awareness of safety measures for example.

Platforms may exhibit any of these features to different extents, while some have aspects combine or reject some of these features and thus can blur the lines between private and public further. To illustrate, many social media platforms allow for the creation of groups or specific discussion threads.

On Facebook, for example, users can create discussion groups where several discussions can be observed in a stream like manner, yet the posts may not be public and are limited to the topic or theme at hand. Some groups require screening prior to admission while other are free for anyone to join. Regardless, the selected audience of the group may induce different understandings and expectations of privacy which does not make the public or private status of the group as clear cut. As a different example, Reddit mainly hosts discussion threads on specified topics instead of a general non-specific stream of content. In this way all or almost all discussions are public, however, they are already siloed into particular topics instead of free-standing posts.

8.4.2. General Social Media Platforms

The most widely used social media platforms, such as Facebook, Instagram, TikTok and X are so-called general social media platforms. These platforms are not catering for a particular theme or topic. It was estimated that in 2022 around 1.7 million pieces of content were created on Facebook, 2.4 million on Snapchat, and 347 000 on Twitter every minute of every day (Domo, 2022). As such these platforms contain enormous amounts of data, of which some is created by patients, HCPs and caregivers. It should be noted that given the wealth of posts on all topics, there is a significant amount of data not relevant to regulatory use-cases or patient data, and a significant portion of fabricated data which must be filtered out. Relevant patient data is present in patients' user profiles as well as in public user posts on the stream related to topics of interest. Additionally relevant data may be present in private content such as closed groups or direct messages; however, these are inaccessible.

8.4.3. Patient Focused Platforms

A plethora of social media platforms targeted for patients are also widely used today. PatientsLikeMe hosts around 830 000 users on their platforms while more than 230 000 people use HealthUnlocked daily (*HealthUnlocked* | *The Social Network for Health*, n.d.; *PatientsLikeMe*, n.d.). These platforms allow for users to find peers and access and create content by other users with the same disease or who follow the same treatments. As such, the data present on these platforms is highly relevant and categorized to various therapeutic areas. Moreover, discussions on these platforms are more likely to include more complete data in terms of drug doses, medical history, and clinical care than discussions on general social media platforms overall. Some patient focused platforms even collect medical data in a standardised format through questionnaires which makes data analysis more feasible.

9. References

Alessa, A., & Faezipour, M. (2018). A review of influenza detection and prediction through social networking sites. *Theoretical Biology and Medical Modelling*, *15*(1), 2. https://doi.org/10.1186/s12976-017-0074-5

Al-Hamid, A., Tudor, C., & Assi, S. (2024). Exploring profile, effects and toxicity of novel synthetic opioids and classical opioids via Twitter: A qualitative study. *Emerging Trends in Drugs, Addictions, and Health*, *4*, 100139. https://doi.org/10.1016/j.etdah.2023.100139

Almeida, D., Umuhire, D., Gonzalez-Quevedo, R., António, A., Burgos, J. G., Verpillat, P., Bere, N., Sepodes, B., & Torre, C. (2024). Leveraging patient experience data to guide medicines development, regulation, access decisions and clinical care in the EU. *Frontiers in Medicine*, *11*. https://doi.org/10.3389/fmed.2024.1408636

Altarifi, A. A., Baker, M., Abedal-Kareem, K., Abu-Ishqair, A., AbuMelhim, Z., Abu Shetayyah, S., & Almhdawi, K. A. (2024). Knowledge and Attitude of the General Public Toward Palliative Care in Jordan: A Cross-Sectional Study. *American Journal of Hospice and Palliative Medicine®*, 10499091241231780. https://doi.org/10.1177/10499091241231781

AMARS in the EU. (n.d.). Retrieved June 18, 2024, from https://transparency.x.com/en/reports/amars-in-the-eu.html

Anderson, A., Pesa, J., Choudhry, Z., Brethenoux, C., Furey, P., Jackson, L., Valleta, L. G., Quijano, L. G., & Lorenzo, A. (2024). Patient perceptions of disease burden and treatment of myasthenia gravis based on sentiment analysis of digital conversations. *Scientific Reports*, *14*(1), 7271. https://doi.org/10.1038/s41598-024-57825-1

Anwar, A., Malik, M., Raees, V., & Anwar, A. (2020). Role of Mass Media and Public Health Communications in the COVID-19 Pandemic. *Cureus*. https://doi.org/10.7759/cureus.10453

Athanasiou, M., Fragkozidis, G., Zarkogianni, K., & Nikita, K. S. (2023). Long Short-term Memory-Based Prediction of the Spread of Influenza-Like Illness Leveraging Surveillance, Weather, and Twitter Data: Model Development and Validation. *Journal of Medical Internet Research*, *25*, e42519. https://doi.org/10.2196/42519

Bakker, E., Plueschke, K., Jonker, C. J., Kurz, X., Starokozhko, V., & Mol, P. G. M. (2023). Contribution of Real-World Evidence in European Medicines Agency's Regulatory Decision Making. *Clinical Pharmacology & Therapeutics*, *113*(1), 135–151. https://doi.org/10.1002/cpt.2766

Basch, C. H., Narayanan, S., Tang, H., Fera, J., & Basch, C. E. (2023). Descriptive analysis of TikTok videos posted under the hashtag #Ozempic. *Journal of Medicine, Surgery, and Public Health*, *1*, 100013. https://doi.org/10.1016/j.glmedi.2023.100013

Bayer, J. B., Triệu, P., & Ellison, N. B. (2020). Social Media Elements, Ecologies, and Effects. *Annual Review of Psychology*, *71*(Volume 71, 2020), 471–497. https://doi.org/10.1146/annurev-psych-010419-050944

Blaser, D. A., Eaneff, S., Loudon-Griffiths, J., Roberts, S., Phan, P., Wicks, P., & Weatherall, J. (2017). Comparison of rates of nausea side effects for prescription medications from an online patient community versus medication labels: An exploratory analysis. *AAPS Open*, *3*(1), 10. https://doi.org/10.1186/s41120-017-0020-y

Brosch, S., de Ferran, A.-M., Newbould, V., Farkas, D., Lengsavath, M., & Tregunno, P. (2019). Establishing a Framework for the Use of Social Media in Pharmacovigilance in Europe. *Drug Safety*, *42*(8), 921–930. https://doi.org/10.1007/s40264-019-00811-8

Callard, F., & Perego, E. (2021). How and why patients made Long Covid. *Social Science & Medicine*, *268*, 113426. https://doi.org/10.1016/j.socscimed.2020.113426

Cano-Marin, E., Ribeiro-Soriano, D., Mardani, A., & Blanco Gonzalez-Tejero, C. (2023). Exploring the challenges of the COVID-19 vaccine supply chain using social media analytics: A global perspective. *Sustainable Technology and Entrepreneurship*, *2*(3), 100047. https://doi.org/10.1016/j.stae.2023.100047

Chen, J., & Wang, Y. (2021). Social Media Use for Health Purposes: Systematic Review. *Journal of Medical Internet Research*, *23*(5), e17917. https://doi.org/10.2196/17917

Chiauzzi, E., & Wicks, P. (2019). Digital Trespass: Ethical and Terms-of-Use Violations by Researchers Accessing Data From an Online Patient Community. *Journal of Medical Internet Research*, *21*(2), e11985. https://doi.org/10.2196/11985

Coman, D. L., Chard, M. P., Desautels, L., Lutz, B. J., & Minns, L. A. (2024). What do spouse primary caregivers of patients with glioblastoma want medical providers to know? A qualitative thematic reflexive analysis of letters written by primary caregivers from a secret Facebook support group. *BMJ Open*, *14*(3), e081783. https://doi.org/10.1136/bmjopen-2023-081783

DasMahapatra, P., Raja, P., Gilbert, J., & Wicks, P. (2017). Clinical trials from the patient perspective: Survey in an online patient community. *BMC Health Services Research*, *17*(1), 166. https://doi.org/10.1186/s12913-017-2090-x

Digital Services Act (DSA): Information for EU users. (2024, February 16). Reddit Help. https://support.reddithelp.com/hc/en-us/articles/23595536875796-Digital-Services-Act-DSA-Information-for-EU-users

Dirkson, A., Verberne, S., Kraaij, W., van Oortmerssen, G., & Gelderblom, H. (2022). Automated gathering of real-world data from online patient forums can complement pharmacovigilance for rare cancers. *Scientific Reports*, *12*(1), 10317. https://doi.org/10.1038/s41598-022-13894-8

Domo. (2022). *Data Never Sleeps 10.0*. https://web-assets.domo.com/miyagi/images/product/product-feature-22-data-never-sleeps-10.png

ECDC. (n.d.). *Internal procedure on manual monitoring of social media for epidemic intelligence*.

ECDC. (2024, January 30). *Information on social media monitoring for epidemic intelligence purposes*. https://www.ecdc.europa.eu/en/information-social-media-monitoring-epidemic-intelligence-purposes

Edinger, A., Valdez, D., Walsh-Buhi, E., Trueblood, J. S., Lorenzo-Luaces, L., Rutter, L. A., & Bollen, J. (2023). Misinformation and Public Health Messaging in the Early Stages of the Mpox Outbreak: Mapping the Twitter Narrative With Deep Learning. *Journal of Medical Internet Research*, *25*(1), e43841. https://doi.org/10.2196/43841

Eichler, G. S., Cochin, E., Han, J., Hu, S., Vaughan, T. E., Wicks, P., Barr, C., & Devenport, J. (2016). Exploring Concordance of Patient-Reported Information on PatientsLikeMe and Medical Claims Data at the Patient Level. *Journal of Medical Internet Research*, *18*(5), e5130. https://doi.org/10.2196/jmir.5130

Eindor-Abarbanel, A., Pinchevski, N., Shalem, T., Agajany, N., Ophir, N., Weiss, B., Broide, E., & Richter, V. (2024). Parental perspectives on pediatric inflammatory bowel disease: Unraveling concerns, and study participation willingness. *Journal of Pediatric Gastroenterology and Nutrition*, *78*(4), 862–870. https://doi.org/10.1002/jpn3.12172

Emam, K. E., Jonker, E., Arbuckle, L., & Malin, B. (2011). A Systematic Review of Re-Identification Attacks on Health Data. *PLOS ONE*, *6*(12), e28071. https://doi.org/10.1371/journal.pone.0028071

European Medicines Agency. (n.d.). *Modafinil article 31 referral annex i ii iii iv*. https://www.ema.europa.eu/en/documents/referral/modafinil-article-31-referral-annex-i-ii-iii-iv_en.pdf

European Medicines Agency. (2020). *European medicines agencies network strategy to 2025*. https://www.ema.europa.eu/en/documents/report/european-union-medicines-agencies-network-strategy-2025-protecting-public-health-time-rapid-change_en.pdf

European Medicines Agency. (2022a). *EMA's Regulatory Science Strategy to 2025, Mid-point achievements to end 2022*. https://www.ema.europa.eu/en/documents/report/emas-regulatory-science-strategy-2025-mid-point-achievements-end-2022_en.pdf

European Medicines Agency. (2022b, September 1). *Good Practice Guide for the use of the Metadata Catalogue of Real-World Data Sources*. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/good-practice-guide-use-metadata-catalogue-real-world-data-sources_en.pdf

European Medicines Agency. (2023). *Real-world evidence framework to support EU regulatory decision-making: Report on the experience gained with regulator-led studies from September 2021 to February 2023*. https://www.ema.europa.eu/en/documents/report/real-world-evidence-framework-support-eu-regulatory-decision-making-report-experience-gained-regulator-led-studies-september-2021-february-2023_en.pdf

European Medicines Agency. (2024, April 15). *Reflection paper use real-world data non-interventional studies generate real-world evidence*. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-real-world-data-non-interventional-studies-generate-real-world-evidence_en.pdf

Facebook ad ban may squelch medical research recruitment. (2022, January 2). POLITICO. https://www.politico.com/news/2022/01/02/facebook-ad-ban-medical-research-recruitment-526275

Fjallhed, A., Lufkens, M., & Sandre, A. (2023). New Trends in Digital Diplomacy: The Rise of TikTok and the Geopolitics of Algorithmic Governance. In *The Oxford Handbook of Digital Diplomacy*. Oxford University Press.

Flynn, R., Plueschke, K., Quinten, C., Strassmann, V., Duijnhoven, R. G., Gordillo-Marañon, M., Rueckbeil, M., Cohet, C., & Kurz, X. (2022). Marketing Authorization Applications Made to the European Medicines Agency in 2018–2019: What was the Contribution of Real-World Evidence? *Clinical Pharmacology and Therapeutics*, *111*(1), 90–97. https://doi.org/10.1002/cpt.2461

Gao, Y., Duan, W., & Rui, H. (2022). Does social media accelerate product recalls? Evidence from the pharmaceutical industry. *School of Computing and Information Systems at Institutional Knowledge at Singapore Management University*, *33*(Information Systems Research).

Gelder, M. M. H. J. van, Rog, A., Bredie, S. J. H., Kievit, W., Nordeng, H., & Belt, T. H. van de. (2019). Social media monitoring on the perceived safety of medication use during pregnancy: A case study from the Netherlands. *British Journal of Clinical Pharmacology*, *85*(11), 2580–2590. https://doi.org/10.1111/bcp.14083

Gorodnichenko, Y., Pham, T., & Talavera, O. (2021). Social media, sentiment and public opinions: Evidence from #Brexit and #USElection. *European Economic Review*, *136*, 103772. https://doi.org/10.1016/j.euroecorev.2021.103772

Han, S. H., Safeek, R., Ockerman, K., Trieu, N., Mars, P., Klenke, A., Furnas, H., & Sorice-Virk, S. (2024). Public Interest in the Off-Label Use of Glucagon-like Peptide 1 Agonists (Ozempic) for Cosmetic Weight Loss: A Google Trends Analysis. *Aesthetic Surgery Journal*, *44*(1), 60–67. https://doi.org/10.1093/asj/sjad211

Hayawi, K., Saha, S., Masud, M. M., Mathew, S. S., & Kaosar, M. (2023). Social media bot detection with deep learning methods: A systematic review. *Neural Computing and Applications*, *35*(12), 8903–8918. https://doi.org/10.1007/s00521-023-08352-z

HealthUnlocked | The social network for health. (n.d.). HealthUnlocked. Retrieved May 30, 2024, from https://healthunlocked.com

HMA/EMA joint Big Data Steering Group. (2023). *Big Data Workplan 2023-2025*. HMA / EMA. https://www.ema.europa.eu/en/documents/work-programme/workplan-2023-2025-hma-ema-joint-big-data-steering-group_en.pdf

Hossain, T., Logan IV, R. L., Ugarte, A., Matsubara, Y., Young, S., & Singh, S. (2020). COVIDLies: Detecting COVID-19 Misinformation on Social Media. In K. Verspoor, K. B. Cohen, M. Conway, B. de Bruijn, M. Dredze, R. Mihalcea, & B. Wallace (Eds.), *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.nlpcovid19-2.11

Hua, Y., Jiang, H., Lin, S., Yang, J., Plasek, J. M., Bates, D. W., & Zhou, L. (2022). Using Twitter data to understand public perceptions of approved versus off-label use for COVID-19-related medications. *Journal of the American Medical Informatics Association : JAMIA*, *29*(10), 1668–1678. https://doi.org/10.1093/jamia/ocac114

Jain, P., Zaher, Z., & Mazid, I. (2020). Opioids on Twitter: A Content Analysis of Conversations regarding Prescription Drugs on Social Media and Implications for Message Design. *Journal of Health Communication*, *25*(1), 74–81. https://doi.org/10.1080/10810730.2019.1707911

Joshi, G., Srivastava, A., Yagnik, B., Hasan, M., Saiyed, Z., Gabralla, L. A., Abraham, A., Walambe, R., & Kotecha, K. (2023). Explainable Misinformation Detection Across Multiple Social Media Platforms. *IEEE Access*, *11*, 23634–23646. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3251892

Klein, A. Z., Banda, J. M., Guo, Y., Schmidt, A. L., Xu, D., Flores Amaro, I., Rodriguez-Esteban, R., Sarker, A., & Gonzalez-Hernandez, G. (2024). Overview of the 8th Social Media Mining for Health Applications (#SMM4H) shared tasks at the AMIA 2023 Annual Symposium. *Journal of the American Medical Informatics Association*, *31*(4), 991–996. https://doi.org/10.1093/jamia/ocae010

Lamsal, R., Harwood, A., & Read, M. R. (2022). Twitter conversations predict the daily confirmed COVID-19 cases. *Applied Soft Computing*, *129*, 109603. https://doi.org/10.1016/j.asoc.2022.109603

Lee, J.-Y., Lee, Y.-S., Kim, D. H., Lee, H. S., Yang, B. R., & Kim, M. G. (2021). The Use of Social Media in Detecting Drug Safety–Related New Black Box Warnings, Labeling Changes, or Withdrawals: Scoping Review. *JMIR Public Health and Surveillance*, *7*(6), e30137. https://doi.org/10.2196/30137

Li, L., Gao, L., Zhou, J., Ma, Z., Choy, D. F., & Hall, M. A. (2021). *Can Social Media Data Be Utilized to Enhance Early Warning: Retrospective Analysis of the U.S. Covid-19 Pandemic* (p. 2021.04.11.21255285). medRxiv. https://doi.org/10.1101/2021.04.11.21255285

Li, Y., Salmasian, H., Harpaz, R., Chase, H., & Friedman, C. (2011). Determining the Reasons for Medication Prescriptions in the EHR using Knowledge and Natural Language Processing. *AMIA Annual Symposium Proceedings*, *2011*, 768–776.

Lopreite, M., Panzarasa, P., Puliga, M., & Riccaboni, M. (2021). Early warnings of COVID-19 outbreaks across Europe from social media. *Scientific Reports*, *11*(1), 2147. https://doi.org/10.1038/s41598-021-81333-1

Macdonald, J. C., Isom, D. C., Evans, D. D., & Page, K. J. (2021). Digital Innovation in Medicinal Product Regulatory Submission, Review, and Approvals to Create a Dynamic Regulatory

Ecosystem—Are We Ready for a Revolution? *Frontiers in Medicine*, *8*, 660808. https://doi.org/10.3389/fmed.2021.660808

Marshall, S. A., Yang, C. C., Ping, Q., Zhao, M., Avis, N. E., & Ip, E. H. (2016). Symptom clusters in women with breast cancer: An analysis of data from social media and a research study. *Quality of Life Research*, *25*(3), 547–557. https://doi.org/10.1007/s11136-015-1156-7

Maskell, C. (2022, August 11). *Social Media and Real-World Evidence Community of Experts*. Pistoia Alliance. https://www.pistoiaalliance.org/projects/current-projects/social-media-rwe-community/

Mohammed, F., Al-Kumaim, N. H., Alzahrani, A. I., & Fazea, Y. (2023). The Impact of Social Media Shared Health Content on Protective Behavior against COVID-19. *International Journal of Environmental Research and Public Health*, *20*(3), 1775. https://doi.org/10.3390/ijerph20031775

Mühlhoff, R., & Willem, T. (2023). Social media advertising for clinical studies: Ethical and data protection implications of online targeting. *Big Data & Society*, *10*(1), 20539517231156130. https://doi.org/10.1177/20539517231156127

Mullin, A., Scott, M., Vaccaro, G., Floresta, G., Arillotta, D., Catalani, V., Corkery, J. M., Stair, J. L., Schifano, F., & Guirguis, A. (2024). Benzodiazepine Boom: Tracking Etizolam, Pyrazolam, and Flubromazepam from Pre-UK Psychoactive Act 2016 to Present Using Analytical and Social Listening Techniques. *Pharmacy*, *12*(1), Art. 1. https://doi.org/10.3390/pharmacy12010013

Nasralah, T., El-Gayar, O., & Wang, Y. (2020). Social Media Text Mining Framework for Drug Abuse: Development and Validation Study With an Opioid Crisis Case Analysis. *Journal of Medical Internet Research*, *22*(8), e18350. https://doi.org/10.2196/18350

Orabi, M., Mouheb, D., Al Aghbari, Z., & Kamel, I. (2020). Detection of Bots in Social Media: A Systematic Review. *Information Processing & Management*, *57*(4), 102250. https://doi.org/10.1016/j.ipm.2020.102250

Ostler, C., Dickinson, A., Metcalf, C., & Donovan-Hall, M. (2024). Exploring the patient experience and perspectives of taking part in outcome measurement during lower limb prosthetic rehabilitation: A qualitative study. *Disability and Rehabilitation*, *0*(0), 1–11. https://doi.org/10.1080/09638288.2024.2307384

Oyebode, O., & Orji, R. (2023). Identifying adverse drug reactions from patient reviews on social media using natural language processing. *Health Informatics Journal*, *29*(1), 14604582221136712. https://doi.org/10.1177/14604582221136712

Pacific, T. L. R. H.-W. (2023). Where are the drugs? The scarcity of medications in the Western Pacific. *The Lancet Regional Health: Western Pacific*, *31*. https://doi.org/10.1016/j.lanwpc.2023.100728

Park, S., Choi, S. H., Song, Y.-K., & Kwon, J.-W. (2022). Comparison of Online Patient Reviews and National Pharmacovigilance Data for Tramadol-Related Adverse Events: Comparative Observational Study. *JMIR Public Health and Surveillance*, *8*(1), e33311. https://doi.org/10.2196/33311

PatientsLikeMe. (n.d.). PatientsLikeMe. Retrieved May 30, 2024, from https://www.patientslikeme.com/

Puspitasari, I., & Firdauzy, A. (2019). Characterizing Consumer Behavior in Leveraging Social Media for E-Patient and Health-Related Activities. *International Journal of Environmental Research and Public Health*, *16*(18), Art. 18. https://doi.org/10.3390/ijerph16183348

Reynolds, T. L., Ali, N., McGregor, E., O'Brien, T., Longhurst, C., Rosenberg, A. L., Rudkin, S. E., & Zheng, K. (2018). Understanding Patient Questions about their Medical Records in an Online Health Forum: Opportunity for Patient Portal Design. *AMIA Annual Symposium Proceedings*, *2017*, 1468–1477.

Rezaallah, B., Lewis, D. J., Pierce, C., Zeilhofer, H.-F., & Berg, B.-I. (2019). Social Media Surveillance of Multiple Sclerosis Medications Used During Pregnancy and Breastfeeding: Content Analysis. *Journal of Medical Internet Research*, *21*(8), e13003. https://doi.org/10.2196/13003

Roborel de Climens, A., Pain, E., Boss, A., & Shaunik, A. (2020). Understanding Reasons for Treatment Discontinuation, Attitudes and Education Needs Among People Who Discontinue Type 2 Diabetes Treatment: Results from an Online Patient Survey in the USA and UK. *Diabetes Therapy*, *11*(8), 1873–1881. https://doi.org/10.1007/s13300-020-00843-9

Saha, K., Sugar, B., Torous, J., Abrahao, B., Kıcıman, E., & Choudhury, M. D. (2019). A Social Media Study on the Effects of Psychiatric Medication Use. *Proceedings of the International AAAI Conference on Web and Social Media*, *13*, 440–451. https://doi.org/10.1609/icwsm.v13i01.3242

Sarker, A., DeRoos, A., & Perrone, J. (2020). Mining social media for prescription medication abuse monitoring: A review and proposal for a data-centric framework. *Journal of the American Medical Informatics Association*, *27*(2), 315–329. https://doi.org/10.1093/jamia/ocz162

Song, Y.-K., Song, J., Kim, K., & Kwon, J.-W. (2022). Potential Adverse Events Reported With the Janus Kinase Inhibitors Approved for the Treatment of Rheumatoid Arthritis Using Spontaneous Reports and Online Patient Reviews. *Frontiers in Pharmacology*, *12*. https://doi.org/10.3389/fphar.2021.792877

Sunkureddi, P., Doogan, S., Heid, J., Benosman, S., Ogdie, A., Martin, L., & Palmer, J. B. (2018). Evaluation of Self-reported Patient Experiences: Insights from Digital Patient Communities in Psoriatic Arthritis. *The Journal of Rheumatology*, *45*(5), 638–647. https://doi.org/10.3899/jrheum.170500

van Stekelenborg, J., Ellenius, J., Maskell, S., Bergvall, T., Caster, O., Dasgupta, N., Dietrich, J., Gama, S., Lewis, D., Newbould, V., Brosch, S., Pierce, C. E., Powell, G., Ptaszyńska-Neophytou, A., Wiśniewski, A. F. Z., Tregunno, P., Norén, G. N., & Pirmohamed, M. (2019). Recommendations for the Use of Social Media in Pharmacovigilance: Lessons from IMI WEB-RADR. *Drug Safety*, *42*(12), 1393–1407. https://doi.org/10.1007/s40264-019-00858-7

Velardi, P., Stilo, G., Tozzi, A. E., & Gesualdo, F. (2014). Twitter mining for fine-grained syndromic surveillance. *Artificial Intelligence in Medicine*, *61*(3), 153–163. https://doi.org/10.1016/j.artmed.2014.01.002

Walsh, J., Cave, J., & Griffiths, F. (2021). Spontaneously Generated Online Patient Experience of Modafinil: A Qualitative and NLP Analysis. *Frontiers in Digital Health*, *3*. https://doi.org/10.3389/fdgth.2021.598431

Walsh, J., Dwumfour, C., Cave, J., & Griffiths, F. (2022). Spontaneously generated online patient experience data - how and why is it being used in health research: An umbrella scoping review. *BMC Medical Research Methodology*, *22*(1), 139. https://doi.org/10.1186/s12874-022-01610-z

Wang, W., Wang, Y., Zhang, X., Jia, X., Li, Y., & Dang, S. (2020). Using WeChat, a Chinese Social Media App, for Early Detection of the COVID-19 Outbreak in December 2019: Retrospective Study. *JMIR MHealth and UHealth*, *8*(10), e19589. https://doi.org/10.2196/19589

Ware, O. D., Garcia-Romeu, A., Zamarripa, C. A., Hughes, T., Wager, L., & Spindle, T. (2024). Codeine and promethazine: Exploratory study on "lean" or "sizzurp" using national survey data and an online forum. *PLOS ONE*, *19*(3), e0301024. https://doi.org/10.1371/journal.pone.0301024

Wessel, D., & Pogrebnyakov, N. (2024). Using Social Media as a Source of Real-World Data for Pharmaceutical Drug Development and Regulatory Decision Making. *Drug Safety*, *47*(5), 495–511. https://doi.org/10.1007/s40264-024-01409-5

White Hughto, J. M., Pachankis, J. E., Eldahan, A. I., & Keene, D. E. (2017). "You Can't Just Walk Down the Street and Meet Someone": The Intersection of Social–Sexual Networking Technology, Stigma, and Health Among Gay and Bisexual Men in the Small City. *American Journal of Men's Health*, *11*(3), 726–736. https://doi.org/10.1177/1557988316679563

Willis, E., & Delbaere, M. (2022). Patient Influencers: The Next Frontier in Direct-to-Consumer Pharmaceutical Marketing. *Journal of Medical Internet Research*, *24*(3), e29422. https://doi.org/10.2196/29422

Young, S. D., Lee, S.-J., Perez, H., Gill, N., Gelberg, L., & Heinzerling, K. (2020). Social media as an emerging tool for reducing prescription opioid misuse risk factors. *Heliyon*, *6*(3). https://doi.org/10.1016/j.heliyon.2020.e03471

Zhang, Y., Lyu, H., Liu, Y., Zhang, X., Wang, Y., & Luo, J. (2021). Monitoring Depression Trends on Twitter During the COVID-19 Pandemic: Observational Study. *JMIR Infodemiology*, *1*(1), e26769. https://doi.org/10.2196/26769

Ziauddeen, N., Gurdasani, D., O'Hara, M. E., Hastie, C., Roderick, P., Yao, G., & Alwan, N. A. (2022). Characteristics and impact of Long Covid: Findings from an online survey. *PLOS ONE*, *17*(3), e0264331. https://doi.org/10.1371/journal.pone.0264331

Zunic, A., Corcoran, P., & Spasic, I. (2020). Sentiment Analysis in Health and Well-Being: Systematic Review. *JMIR Medical Informatics*, *8*(1), e16023. https://doi.org/10.2196/16023