



OPEN ACCESS

EDITED BY

Carla Matos Torre,
University of Lisbon, Portugal

REVIEWED BY

Prathamesh Karmalkar,
Merck Group, India
Carlos Alves,
University of Coimbra, Portugal

*CORRESPONDENCE

Philipp Cimiano
✉ cimiano@cit-ec.uni-bielefeld.de

RECEIVED 08 August 2023

ACCEPTED 12 February 2024

PUBLISHED 06 March 2024

CITATION

Cimiano P, Collins B, De Vuono MC, Escudier T, Gottowik J, Hartung M, Leddin M, Neupane B, Rodriguez-Esteban R, Schmidt AL, Starke-Knäusel C, Voorhaar M and Wieckowski K (2024) Patient listening on social media for patient-focused drug development: a synthesis of considerations from patients, industry and regulators.

Front. Med. 11:1274688.

doi: 10.3389/fmed.2024.1274688

COPYRIGHT

© 2024 Cimiano, Collins, De Vuono, Escudier, Gottowik, Hartung, Leddin, Neupane, Rodriguez-Esteban, Schmidt, Starke-Knäusel, Voorhaar and Wieckowski. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Patient listening on social media for patient-focused drug development: a synthesis of considerations from patients, industry and regulators

Philipp Cimiano^{1,2*}, Ben Collins³, Maria Carmela De Vuono⁴, Thierry Escudier⁵, Jürgen Gottowik⁶, Matthias Hartung¹, Mathias Leddin⁶, Bikalpa Neupane⁷, Raul Rodriguez-Esteban⁶, Ana Lucia Schmidt⁶, Cornelius Starke-Knäusel¹, Maarten Voorhaar³ and Krzysztof Wieckowski⁸

¹Semalytix GmbH, Bielefeld, Germany, ²CITEC, Bielefeld University, Bielefeld, Germany, ³Boehringer Ingelheim International GmbH, Ingelheim, Germany, ⁴Chiesi Farmaceutici SpA, Parma, Italy, ⁵Pistoia Alliance, Wakefield, MA, United States, ⁶Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd., Basel, Switzerland, ⁷Takeda Pharmaceuticals Co., Ltd., Cambridge, MA, United States, ⁸SageFiber, Ustron, Poland

Patients, life science industry and regulatory authorities are united in their goal to reduce the disease burden of patients by closing remaining unmet needs. Patients have, however, not always been systematically and consistently involved in the drug development process. Recognizing this gap, regulatory bodies worldwide have initiated patient-focused drug development (PFDD) initiatives to foster a more systematic involvement of patients in the drug development process and to ensure that outcomes measured in clinical trials are truly relevant to patients and represent significant improvements to their quality of life. As a source of real-world evidence (RWE), social media has been consistently shown to capture the first-hand, spontaneous and unfiltered disease and treatment experience of patients and is acknowledged as a valid method for generating patient experience data by the Food and Drug Administration (FDA). While social media listening (SML) methods are increasingly applied to many diseases and use cases, a significant piece of uncertainty remains on how evidence derived from social media can be used in the drug development process and how it can impact regulatory decision making, including legal and ethical aspects. In this policy paper, we review the perspectives of three key stakeholder groups on the role of SML in drug development, namely patients, life science companies and regulators. We also carry out a systematic review of current practices and use cases for SML and, in particular, highlight benefits and drawbacks for the use of SML as a way to identify unmet needs of patients. While we find that the stakeholders are strongly aligned regarding the potential of social media for PFDD, we identify key areas in which regulatory guidance is needed to reduce uncertainty regarding the impact of SML as a source of patient experience data that has impact on regulatory decision making.

KEYWORDS

patient-focused drug development, patient experience data, social media listening, real-world evidence, regulatory decision making

1 Introduction

Patients have direct and first-hand experience and knowledge of which symptoms are most burdensome and how the condition affects their daily living, which makes them experts on many aspects of their condition (1). Moreover, they are able to judge their level of unmet need, the gap between the concerns and problems that are addressed by existing treatments and those that remain unsolved. An important goal for drug development is to narrow down this gap and continuously increase the quality of life of the patients. Thus, it has been noted that there is a significant need for more involvement of patients in drug development, in particular in activities related to research prioritization, target product profile development, trial design, regulatory approval, access to medicines, reimbursement, and treatment decisions, which all profit from alignment with the needs of patients (2, 3). For instance, an important measure to make drug development more patient-centric is to ensure that trial outcomes are relevant, appropriate and of importance to patients in real-world settings (2, 4–7). It has been emphasized that this involvement needs to be from early on, ideally already in research prioritization as well as when developing a product strategy, in order to unfold maximal impact (2, 3).

Regrettably, patients have not always been systematically, consistently, and continuously involved in the drug development process (8). This can be seen in many trials that fail to translate into real-world benefits for patients (9).¹ The lack of systematic inclusion of patient voices in drug development is a missed opportunity with respect to the goal of developing treatments that truly improve patients' lives, based on a systematic understanding of the challenges they face in their daily lives and the tradeoffs they are willing to accept. It is thus an important goal to ensure that trials are designed and conducted in such a way that they maximize treatment benefit for patients and address existing unmet needs in their real-world context.

In recognition of this situation, there is a consensus among regulators worldwide that a more systematic approach to patient involvement, known as patient centricity, is needed in order to maximize treatment benefit and facilitate treatment uptake by patients. The Food and Drug Administration (FDA) for instance, has established the Patient-Focused Drug Development (PFDD) initiative (12). With a series of 4 guidance documents, the FDA has sought to provide clarification on how *patient-experience data* can be collected and analyzed to ensure that outcomes of clinical trials are indeed relevant and meaningful from the patient perspective. These guidelines include recommendations on methods for data collection (Guidance 1), and methods for determining what is important to patients (Guidance 2). Guideline 3 on how to select, develop or modify

fit-for-purpose clinical outcomes assessments is available as a draft since June 2022. Guideline 4 on incorporating clinical outcome assessments into endpoints for regulatory decision-making is available as a draft since April 2023. Other agencies have followed with similar initiatives. The European Medicines Agency (EMA) has stressed the importance of patient involvement in drug development as part of its 2025 strategy (13). The Medicines and Healthcare products Regulatory Agency (MHRA) has also published a corresponding strategy for increasing patient involvement in drug development (14) and started a pilot requesting sponsors to submit patient experience data as part of their applications.

The FDA has explicitly mentioned some aspects of patient experience data that are relevant for PFDD:

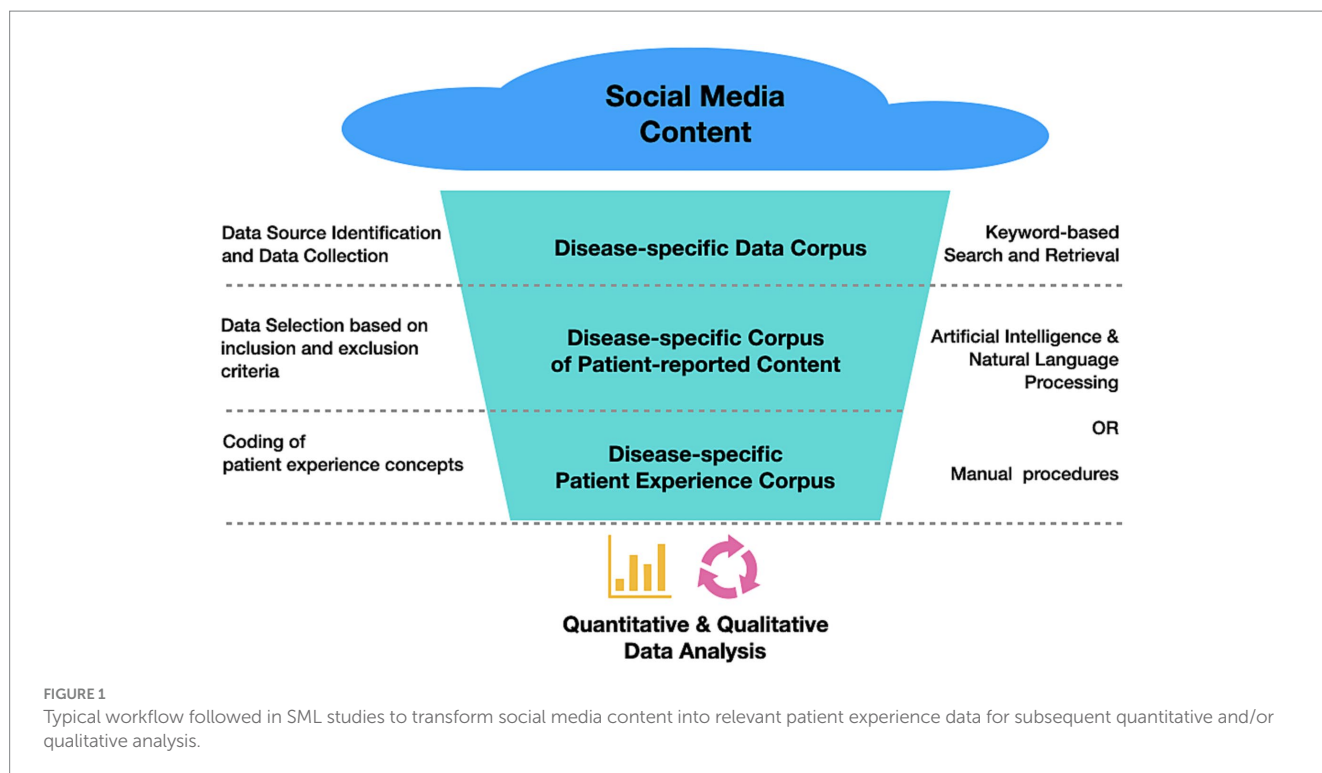
- Impact of the disease and its treatment on the patient (symptoms, chief complaints, burden of living with the disease/condition).
- Patients' perspectives about potential and current treatments (expectation of benefits, tolerance of risks, acceptable tradeoffs).
- Views on unmet medical needs and available treatment options.
- Enhanced understanding of the natural history of the disease or condition.

Because social media has been shown to capture and provide patient insights of relevance to all the dimensions mentioned above (15), it has been specifically mentioned in the guidelines of the FDA as an important source to passively collect patient experience data and to incorporate the voice of patients in drug development both qualitatively and quantitatively. Yet, there remain significant gaps and uncertainties in the understanding of how social media data can be systematically incorporated into drug development and, most importantly, which role it can play in approval processes.

The goal of this paper is to contribute to a holistic perspective on the potential contributions of social media listening (SML) to fostering patient centricity in medical drug development by leveraging online patient experience data. Thus, the paper reviews the current considerations of key stakeholders: patients, pharmaceutical industry and regulators. Based on an analysis of these perspectives, it attempts to provide a synthesis of these, and defines key questions that need to be addressed in future work towards establishing SML as a sound and robust source of patient experience data that impacts regulatory decision making.

The article is structured as follows: section 2 provides a conceptual overview of SML for PFDD from a methodological perspective and a literature survey of published SML studies to shed light on current practice in the field. This also includes a comparison of methodological advantages and challenges of SML as discussed in the literature. Subsequent sections are devoted to the individual perspectives of patients, pharmaceutical industry and regulatory authorities, respectively. The discussion will be primarily guided by (i) exploring to what extent patients consider SML a legitimate and acceptable way to facilitate PFDD (section 3), (ii) the opportunities envisaged by the pharmaceutical industry in applying SML and the legal and ethical framework (section 4), and (iii) current regulatory provisions around using SML for PFDD and involved decision making processes (section 5). Section 6 provides a synthesis of the different stakeholders' perspectives; section 7 concludes the paper with recommendations on an agenda covering six concrete fields for future collaborative action.

¹ For example, in a review of 57 dementia drug trials, Molnar et al. (10) found that less than half (46%) of the trials discussed the clinical significance of their results. Similarly, a systematic review regarding the inclusion of patient-reported outcomes in cardiovascular trials showed that only 23% out of 413 analyzed trials included outcomes of importance to patients. In 40% of these cases, included patient reported outcomes (PROs) were judged to be of little value and 70% of trials were missing data relevant to judge the clinical meaningfulness of results (11).



2 Social media listening for patient-focused drug development

2.1 Methodology

SML refers to a set of observational methods comprising the passive identification, collection and analysis of patient experience data from online data sources, social media in particular. In contrast to other, more established methods such as HRQoL surveys, 1-on-1 interviews, focus groups, etc. these methods are passive in the sense that they neither require nor allow for a direct interaction with patients, who can remain anonymous. Throughout this paper, we will assume a strict definition of “patient” as individual persons who identify themselves in their authored social media posts as being affected by a disease or medical condition (thus excluding caregivers, relatives, healthcare practitioners or other individuals or organizations providing a third-person perspective on a patient’s experience, unless stated otherwise).

Figure 1 presents a conceptual overview of a typical workflow in order to subject social media content to quantitative or qualitative data analysis for the purposes of PFDD comprising at least the following processing steps:

Step 1. Data source identification and data collection: Relevant data sources (e.g., online communities, forums, blogs) for the research questions of interest are identified and retrieved. Relevance criteria are usually bound to diseases or conditions of interest, given that online health communities are typically organized around diseases or disease areas. Consequently, relevant data sources can be identified using keyword-based approaches to search and retrieve disease-related content.

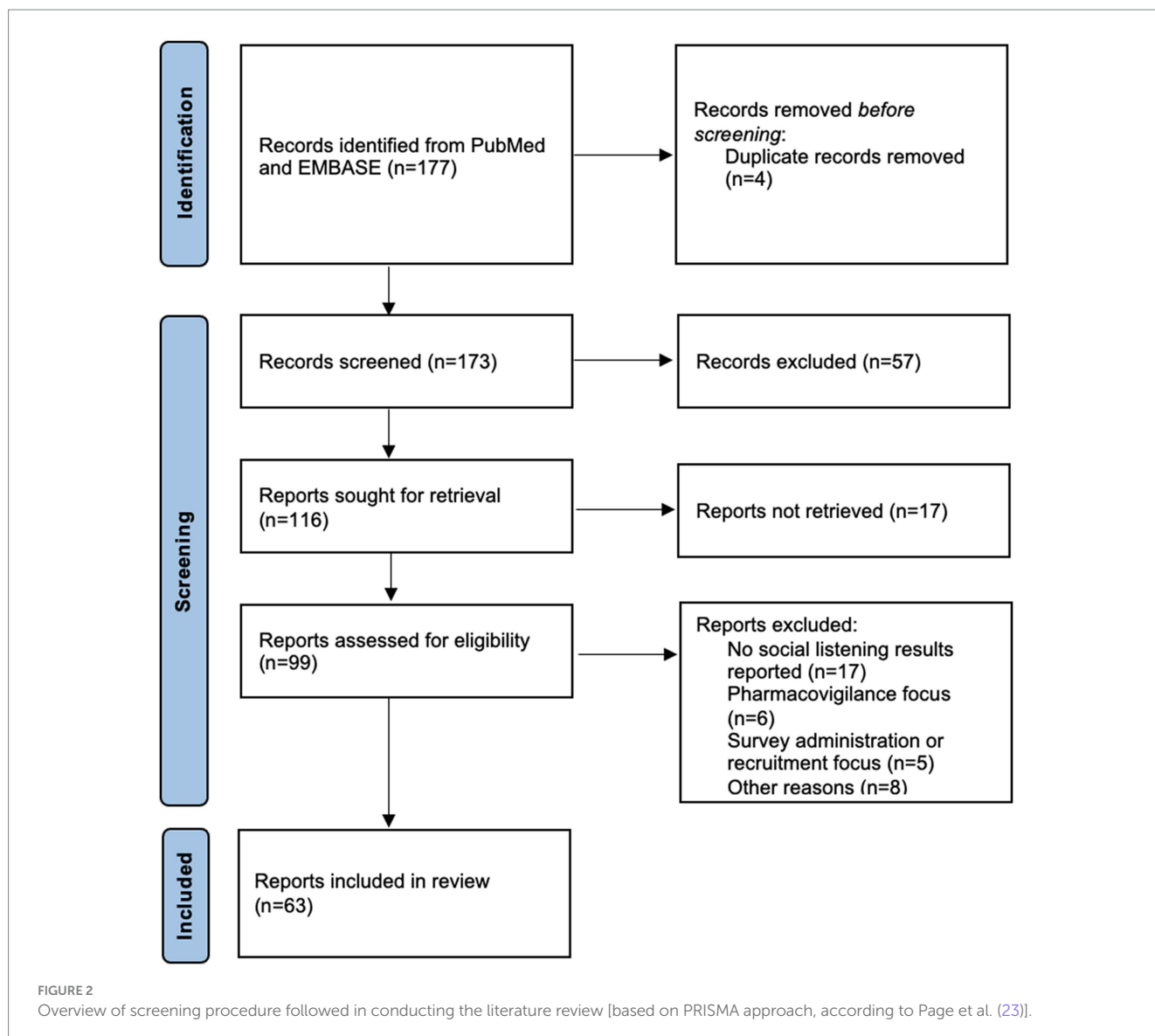
Step 2. Data selection based on inclusion and exclusion criteria: Health-related online content, even if retrieved from carefully

selected sources, usually needs to be filtered according to meaningful criteria such as the perspective of authorship (patients vs. caregivers or other stakeholders) and other inclusion or exclusion criteria of relevance (e.g., demographic variables, disease stages, treatment paradigms, exclusion of certain behaviors, etc.). The content resulting from applying these inclusion and exclusion criteria should be organized in patient-specific data records, while paying close attention to anonymizing all information that could reveal a patient’s identity.

Step 3. Coding of patient experience concepts: Patient data records are processed in order to investigate key concepts of the patient experience (e.g., reports on symptom burden, quality of life impairments, aspects of the treatment experience) and made accessible for subsequent analysis in a quantitative or qualitative setting. This step is usually conducted as a systematic coding (or “labeling”) procedure following clear concept definitions and coding guidelines (16).

It is worth mentioning that, while Step 1 inherently requires algorithmic processing, either algorithmic or manual approaches are feasible in Steps 2 and 3 (at different levels of scalability). However, a growing tendency towards algorithmic procedures capitalizing on natural language processing (NLP) and artificial intelligence (AI) can be recently observed: Consider Spies et al. (17, 18), Staunton et al. (19), Delestre-Levai et al. (20), Freeman et al. (21) or Gries and Fastenau (22) as examples of studies implementing algorithmic solutions to Steps 1–3 mentioned above. In fact, we recommend relying on appropriately conducted algorithmic approaches² to

² For an overview of recommended best practices, please refer to section 4.4 of this article.



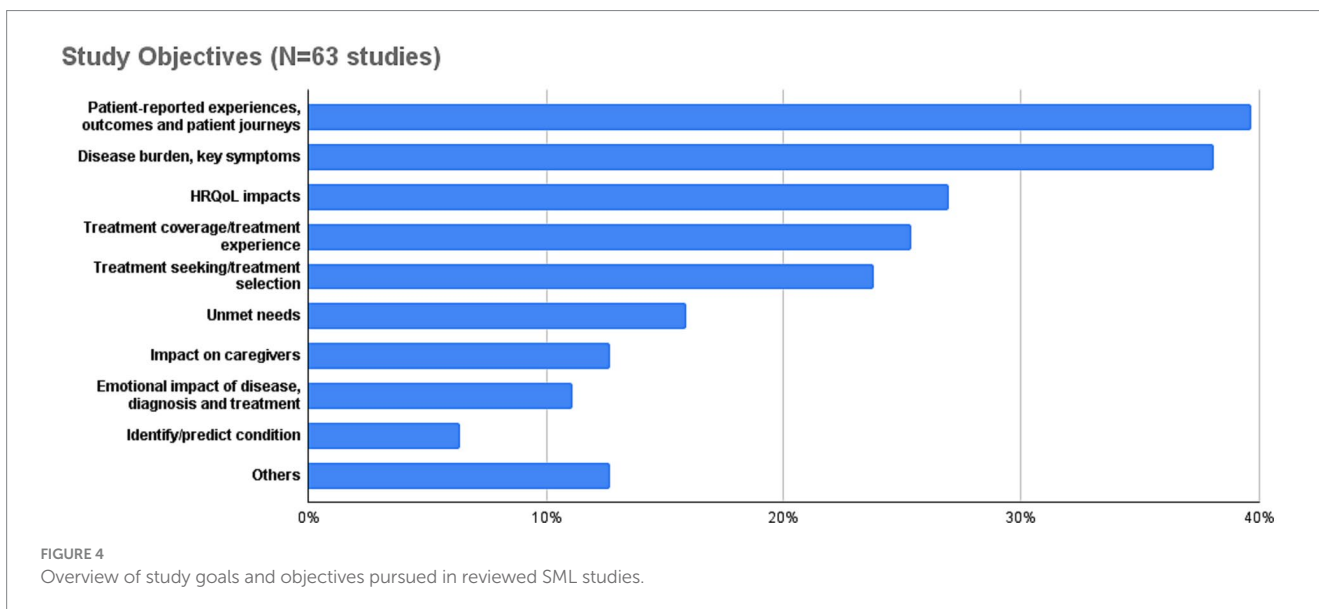
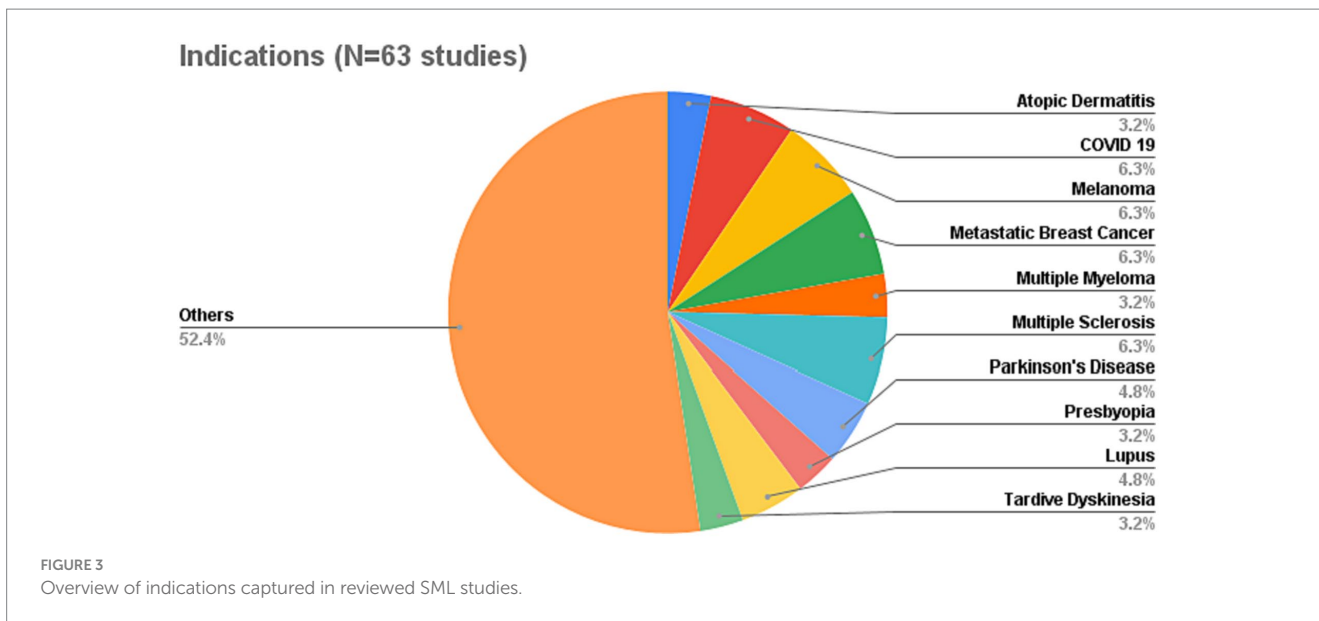
leverage the full potential of SML in collecting a corpus of patient experience data that is most comprehensive and diverse.

2.2 Current practice in SML for PFDD: literature survey

SML is an active area of research with many mature methodologies having been developed to date. In order to summarize the current work and focus in SML as part of this article, a literature review was conducted by querying PubMed and Embase as the most comprehensive subject matter databases at the authors' availability. The search strategy was based on the combination of two queries as documented in Appendix Q. Query 1 was designed to capture a broader set of SML studies, while excluding reports about social media as recruitment, engagement or dissemination channels for studies following different methodological settings. Likewise, studies focusing on different behaviors in social media use were excluded as well. Query 2 was designed to capture SML studies

with an explicitly mentioned focus on PFDD or real-world evidence (RWE) generation. Following this search strategy, 177 publication records from 2015 to February 2023³ were identified and investigated following the PRISMA approach (23) as shown by the flow diagram in Figure 2. As an outcome of the screening procedure, a selection comprising 63 relevant articles with a focus on using SML for PFDD purposes (rather than pharmacovigilance monitoring, patient recruitment, survey administration, or others) was selected for in-depth review, the main findings of which are summarized below. An overview of all articles included can be found in Appendix A.

³ The selected timeframe is motivated by a considerable rise of social media use around 2015 (cf. <https://ourworldindata.org/rise-of-social-media>), which is also reflected in a markedly increased volume of relevant SML studies available from the selected databases after 2015.



2.2.1 Indications

As displayed in Figure 3, SML has been applied to a variety of indications or medical conditions. The most prominent ones include *melanoma* (24–27), *metastatic breast cancer* (28–31), *COVID 19* or *long COVID* (32–35), and *multiple sclerosis* (36–39). Overall, the data set of reviewed studies comprises a total of 40 indications or conditions from a variety of different disease areas (among them cancer, neurological diseases, auto-immune diseases, gastrointestinal diseases, and various others as detailed in Appendix A). Notably, it includes rare diseases, such as *complement 3 glomerulopathy* (40) or *SLC6A1 disorder* (41), which often pose challenges to established research methods due to relatively small patient populations. This variety indicates the suitability of SML as a highly flexible and versatile method to elicit patient experience data across a wide range of disease areas, irrespective of their prevalence.

2.2.2 Study objectives

With respect to main objectives pursued in SML studies, a considerable degree of convergence can be observed, given that most of the objectives mentioned in the reviewed articles can be grouped into a relatively small number of recurring categories (Figure 4): almost 40% of studies have aimed at patient-reported experiences with the potential to inform patient-reported outcomes research or the development of patient journeys. Assessing the burden of disease in terms of key symptoms is of similar relevance for study authors (38%), followed by impacts and impairments on patients’ quality of life (27%). Additionally, more than 20% of studies aim at investigating patients’ treatment experience and related aspects. Note that these focus areas are closely aligned with objectives traditionally pursued by established research instruments in the PFDD space such as surveys, interviews or focus groups. In fact, many authors explicitly emphasize that SML should not be understood as a substitute for existing

research instruments, but as a complement to them, e.g., (20, 22, 27). In fact, many types of other RWE have considerable biases and SML is not unique in that respect.

2.2.3 Data source types

In terms of social media sources, the most widely used ones are X/Twitter (20, 42, 43), online forums (20, 36, 44), Facebook or closed networks (20, 45, 46), and patient communities (25, 47, 48). More recently, studies have also started to use Reddit as a data source, e.g., (36, 49, 50). YouTube and Instagram have also been used in the last 2–3 years, e.g., (33, 36, 46, 47, 51). A comprehensive overview of data source types used in the reviewed studies is shown in Figure 5.

2.2.4 Sponsors

Almost 70% of the SML studies reviewed were conducted with sponsors from the pharmaceutical industry. Novartis (25% of studies reviewed) and Roche (8%) are the most engaged stakeholders in using SML studies for PFDD purposes so far.

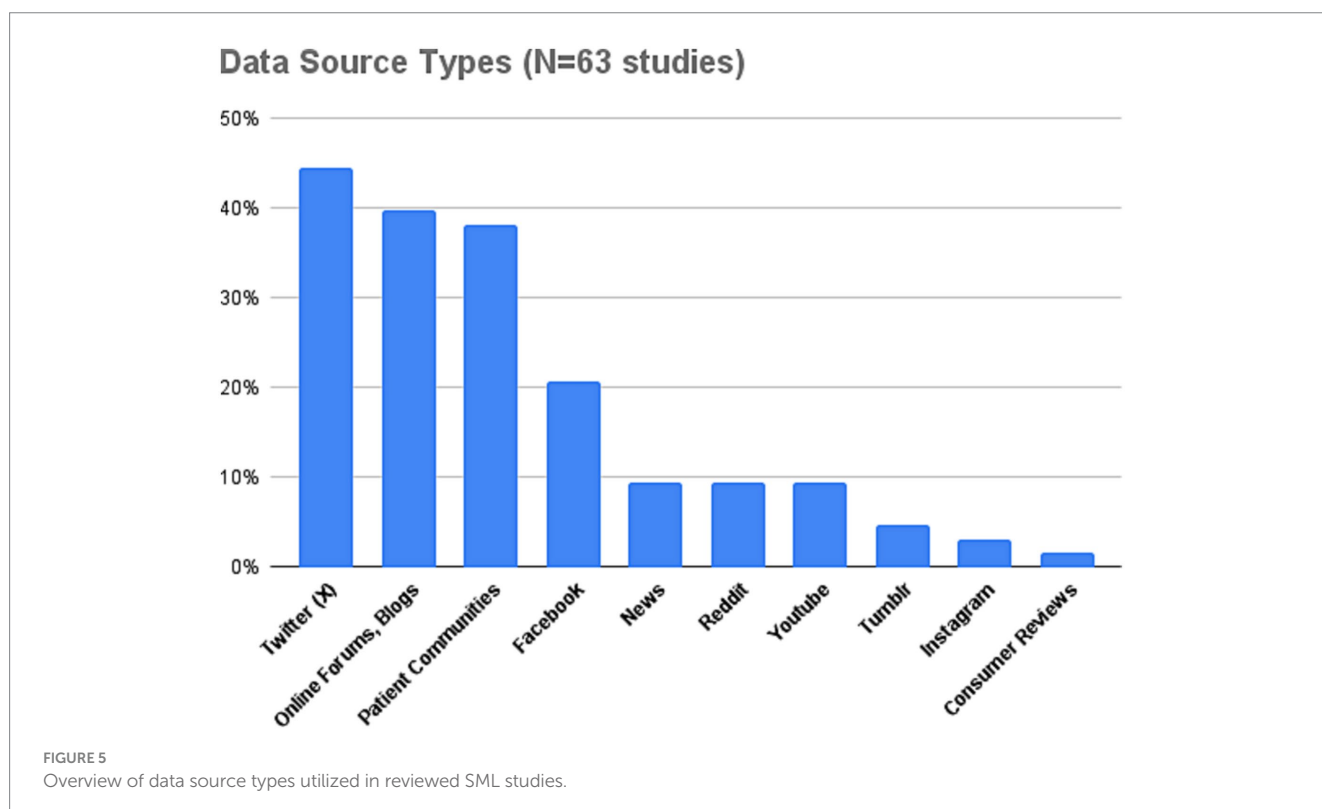
2.3 Benefits and limitations of SML studies

2.3.1 Benefits

The main benefit of SML is that it provides access to the first-person, authentic, spontaneous, unbiased and unfiltered perspective of patients, described in their own words (20, 43, 52). Access to first-person author experiences in patients' own words is particularly valuable for PFDD, as it facilitates conceptualizing the disease from the patients' perspective by using terms and concepts that reflect their understanding.

It has been shown that SML is often able to reproduce the main symptoms of a disease as identified by more traditional methods, in some cases even identifying symptoms that were not captured in interviews (53). Going beyond symptoms, SML has the potential to contribute to a more holistic perspective of the disease burden by helping to understand (i) the factors that reduce the quality of life of patients (22, 54), (ii) the social context into which they are embedded, as well as (iii) their emotional trajectories (20, 22). It can provide background to understand treatment seeking and treatment selection behavior and the main factors involved therein (38, 55, 56). SML allows researchers to track patient perspectives and health trends over time, offering crucial insights into the evolution of diseases and patient experiences (57) and allows to capture evolving insights and trends in real time.

Some authors have emphasized that SML can provide insights about a broader population compared to the more narrow populations considered in clinical trials, as we might be able to capture the perspectives of people that are underrepresented in clinical trials but also the perspectives of family, caregivers, etc. in the broader environment of patients. Thus, in contrast to many other research methods, SML has the unique advantage of allowing to reach patient populations that traditional healthcare approaches might overlook. For instance, individuals suffering from rare diseases, those who lack access to healthcare due to the absence of a healthcare plan, or patients who have discontinued their treatment for a range of reasons—including dissatisfaction with outcomes, severe side effects, or conditions rendering them incapable of treatment (e.g., psychosis)—can be effectively found via social media. It can help in particular to assess conditions in which patient experiences and symptoms may go unreported due to factors such as fear of stigmatization, distrust of healthcare professionals, or personal uncertainty (58). Furthermore,



it can help address diverse and broad populations that might be hard to reach otherwise. Leveraging the expansive reach of social media facilitates the analysis of a diverse and large population sample, hence providing a comprehensive data volume that may be more representative than traditional data collection methods due to the amount of data points and observations related to various aspects of a patient's condition.

From a methodological perspective, SML is an observational, non-interventional method (28, 44, 59) that does not rely on a specific set of questions to be asked to patients, in contrast to more traditional instruments such as interviews, surveys, or focus groups (16). Because it is not restricted to preformulated questions or structure, SML methods avoid problems related to biases that might be created by asking a question in a specific manner (e.g., leading-question bias). For this reason, it has been noted that SML is a suitable method to identify which priorities and outcomes are important to patients from their subjective perception (52). At the same time, as the method simply observes and analyzes the online behavior of patients, it minimizes the burden and requirements on the patient to contribute their views for PFDD (43). Further, SML represents a very cost-effective and scalable method (20) to elicit patient experience data from geographically diverse patient populations (43).

2.3.2 Limitations

An inherent limitation of SML is that, in deciding whether to include a patient in a SML study, we rely on the self-disclosed information of patients to decide whether they meet inclusion and exclusion criteria. The information provided by patients, e.g., on diagnosis, might however be incomplete and inaccurate, leading to uncertainty regarding whether the investigated population really fulfills all inclusion and exclusion criteria (20, 27). In particular, demographic information is not widely and accurately available from social media sources (60). Methodologically, when using SML, it is thus important to take into account the fact that there is uncertainty in terms of the extent to which patients fulfill the inclusion and exclusion criteria. Due to the non-interventional, observational nature of the method, in spite of some comments of patients being unclear or incomplete, there is no possibility for following up or seeking clarification, an inherent limitation of the method (43).

A further important limitation regards the validity of the results obtained with SML and the extent to which they generalize to the target population (20, 27). In particular, data collection from social media might be affected by several types of selection bias. One bias is clearly demographic, as it has been noted that populations active on social media tend to be slightly younger compared to the average of the general population (22, 27). Further, some online communities might emerge as a result of specific unmet needs or experiences, so that the level of unmet need is over-represented in these communities. Finally, as activity on social media follows a power-law distribution (61), online communities might mainly reflect the views of more active users. Finally, groups without access to the internet or those with low digital literacy might be severely underrepresented. In general, user behavior in online communities is fragmented with some users posting actively while others tend to listen to the conversation passively (62). The voices of the more passive users would be under-represented in SML in spite of users consuming social media for healthcare reasons.

Most of the studies conducted in the SML paradigm are particularly sensitive to patients' privacy. Privacy regulations and the existence of closed communities pose considerable challenges to the comprehensive collection of SML data. Often, the data posted and available for public viewing is not allowed to be scraped in a SML project, resulting in selection bias and potentially skewing results. Most published studies are carried out with content that has been made public without any further protection on the Web [e.g., (20, 60, 63)] and apply methods to anonymize patients by removing or pseudo-anonymizing user names [e.g., (20, 22, 55, 60, 63)]. While some studies have obtained explicitly approval by an ethics committee (22, 43, 63), others have mentioned explicitly to be exempt of obtaining approval due to the fact that they used only data that has been explicitly and manifestly made public [see, e.g., (38)]. While it is clearly a best practice in SML to rely only on freely and publicly shared content and applying methods to ensure the non-identifiability of patients, it is an open question if this practice is sufficient to meet the privacy concerns of patients and other stakeholders. A challenge thus continues in adequately addressing privacy concerns and ethical standards (43, 64), which is further discussed in section 4.

2.4 Summary

In summary, SML is an increasingly widely applied method that has been used for a range of different indications and that has matured considerably over the last 5 years. By now, a clear convergence in terms of methods and best practices can be observed.

As any other method for extracting RWE, SML has clear benefits and limitations (as summarized in Table 1) that need to be balanced to pave the way for its wider adoption as an accepted method to capture what matters to patients. The fact that social media is an effective method to obtain the authentic and unfiltered perspective of large numbers of patients comes at the price of a reduced control about the population in terms of demographics and fulfillment of inclusion and exclusion criteria. Given this trade-off, SML can effectively complement existing research instruments that are more established and controllable but might give us a limited perspective of patients' needs.

3 Patient perspective

Our discussion of the patients' own perspective on the prospects of SML for PFDD purposes centers around two fundamental questions: Are patients willing to share their personal, health-related data online via social media? How do patients consider the potential of SML as a method to increase the level of patient involvement in drug development? We discuss these questions in the subsections below, prepending a review of current trends in social media usage for health-related purposes.

3.1 Social media usage for health-related purposes

In a consumer survey among adults in the U.S. to analyze the use of social media for healthcare-related purposes in 2012, the PWC

TABLE 1 Advantages and challenges associated with SML as reported in the literature.

Advantages	Challenges
<ul style="list-style-type: none"> • Identify priorities and outcomes that are important to patients (52) • First-person authenticity, fresh, spontaneous and unbiased patient perspective (63) in their own words (20, 43, 52), unsolicited and unfiltered patient insights • Complement existing data sources and research instruments (20, 22, 64) • Holistic and more complete picture of patient experience (43) • Overcome biases in the design of interviews (43) • Unbiased, non-interventional data (22) • Inform design of other research instruments (22, 65) • Access to social context and emotional journey (20, 22) due to greater disclosure (63) • Follow concepts through full patient journey (22) • Minimize burden and requirements on patients (43) • Facilitate cross-geographic analysis (20, 43) • Represent perspectives of patients who are underrepresented in trials or are difficult to access (63) • Address underrepresented or rare conditions (66) • Access hard-to-reach populations (67) • Analysis of diverse and broad populations (68) • Identification of events vs. longitudinal trends (57) • Cost-effective, scalable approach (20) • Data access in real-time (69) 	<ul style="list-style-type: none"> • Limited generalizability to larger population (20, 27) • Limited access to demographic data (60) • Selection bias (22, 27) • Non-confirmed, self-reported diagnosis (20, 27) • Privacy and ethical concerns (27, 43) • No possibility for following-up or clarification (43) • Negative conversations may be vocalised more often than positive perceptions/experiences (57) • User activity imbalance (62) • Users' tendency to self-diagnose and potentially misdiagnose themselves, especially in the context of mental disorders (70)

Health Research Institute found that 42% of consumers have used social media to access health-related consumer reviews (71); 32% have used social media to view family/friend health experiences, and 29% have sought information related to other patients' experiences with their disease. A more recent report by the Pew Research Center (72) on the use of social media by mobile phone users in emerging markets revealed that 61% of mobile users have looked up information about health and medicine for themselves and their families. In fact, in terms of information seeking activities, looking up health-related information was the top activity.

It has been found in literature studies that patients do not use social media to circumvent healthcare professionals, but rather use it as "a complement to healthcare professional services to fulfill the needs that can not be met by HCPs" (73). One of the crucial reasons for patients to seek support online is their dissatisfaction with professionals' inability to meet their emotional and informational needs. The strive for social support by other patients is thus one of the main reasons for social media use by patients. Smailhodzic et al. have identified the specific types of social support that are sought by patients (73):

- *Emotional support*: referring to support gained through expressions of care and concern.
- *Esteem support*: referring to communication that fosters patients' self-esteem or belief in their ability to handle problems.
- *Information support*: referring to communication that provides needed information.
- *Network support*: communication that affirms patients' belonging to a network or reminds them of support available in their network.

Most importantly, sharing health-related experiences in online communities has an effect of empowerment on patients, increasing their subjective well-being, psychological well-being and leading to increased self-management and control (73).

3.2 Willingness to share personal health-related information for PFDD purposes

When engaging with social media, users are willing to disclose very intimate information (74). In fact, it has been shown that the anonymity of many social media sites increases the level of self-disclosure (75). The most important barriers and problems related to social media use for healthcare purposes are thus privacy issues and issues related to the unreliability of social media (76). In spite of the high sensitivity to privacy, the consumer survey by PWC (71) has shown that more than 30% of respondents would be comfortable having their social media conversations monitored if that data could help identify ways to improve their health. In a study with adults presenting to an academic, urban emergency department, Padrez et al. (77) found out that out of those patients having a social media profile, 71% consented to share their social media data to compare it to their electronic medical records (EMRs). A survey on health topics carried out by Pew (78) corroborates this willingness to share data. In spite of reservations, approximately 70% with a medical condition believe data could potentially be used without their knowledge to deny them healthcare benefits or to deny them job opportunities. More than 90% of interviewees would share their health data if it helps to improve or support research, and approximately 80% would share information with drug companies if it contributes to learning more about the disease or making safer products. These figures convey that a significant share of patients are willing to share their social media data for research purposes if they contribute to improving their condition and those of their peers.

3.3 SML as a facilitator of patient involvement in drug development?

Patient advocacy groups have been increasingly emphasizing the lack of involvement of patients in research design. The International

Neuroendocrine Cancer Alliance (INCA), for instance, is a global alliance made up of 20 patient advocacy groups and research groups in 17 countries from Asia, Asia Pacific, Europe and North America. In a recent survey (79) among (i) patient leaders, (ii) patients and caregivers, as well as (iii) healthcare professionals, INCA identified that over 32% of patient leaders believe that unmet needs of patients are not addressed sufficiently in the current standard of care. Regarding the involvement of patients in research design, 82% of patient leaders and 53% of patient and family think that patients are not sufficiently involved in research design.

Further evidence for the fact that the level of patient-centricity in drug development is regarded as low comes from the AURORA project that has carried out a large survey in order to analyze the ability of sponsors to execute on patient-centricity (80). Between July and November 2017, feedback was gathered from 1,282 participants who chose to take part in an online survey. While all participants agreed on the importance of patient-centricity (95% rated importance over 8 on a scale from 1–10), only 30% had confidence (rating of 8 or higher on a scale from 1–10) in the ability to deliver on patient-centricity. Among patients, this confidence was even lower at just 11.5% (AURORA Project 2016).

Interviews with patient leaders and representatives (2) have highlighted that patients wish to be involved in early phases of drug development and not only towards regulatory approval. The study highlighted that patient representatives emphasized that, while not all patients can directly inform the science behind drug development, the insights they can provide around their perceptions of benefits and risks, relevance of outcomes and overall impact on daily life are invaluable and cannot be provided by any other stakeholder.

As part of the assessment of the status of FDA's PFDD initiative as carried out by the Eastern Research Group (ERG), patient representatives were also interviewed; one important conclusion was that patient representatives felt that a greater attention to psychological aspects of the disease, quality of life and measures of the ability to daily function is needed (81).

3.4 Summary

Social media has an important function for patients, providing informational, emotional and social support to them. The anonymity of social media provides a safe environment in which patients are willing to disclose very sensitive and private experiences.

Patients and patient advocacy groups have emphasized the fact that there are significant unmet needs and demand a higher involvement of patients in research design and evidence generation activities. While privacy is an important barrier, a significant share of patients are willing to share their anonymized data publicly if it has the potential to improve their condition.

Overall, the findings discussed above indicate that social media has the potential of reducing burden and making patients more "involved" in drug development by sharing valuable experiences that can guide drug development initiatives, clinical trial design and patient-reported outcomes research. At the same time, important open questions are still remaining: While a general willingness to share personal health-related information can be attested and goes in hand with a trend of continuously increasing social media usage for

health-related information exchange, more research needs to be dedicated to the specific question as to whether SML in its currently evolved practice (cf. section 2 of this article) is, from the patient perspective, considered as a legitimate and ethical method (see further in section 4) to capture their needs, preferences and priorities.

4 Industry perspective

4.1 Incorporating the patient view in drug development

Pharmaceutical companies strive to develop innovative treatments for different needs and under various biological, technological, medical, economic and social constraints. At the end of the drug development process patients, or their caregivers, make the final decision and express their views in the most direct manner: either by accepting the treatment or by resorting to rejection or non-compliance. This decision is based on the information they have available, which is sometimes conflicting (82), and involves multifaceted considerations and trade-offs that go beyond purely medical aspects such as affordability, trust (83), degree of involvement in the decision (84), risk perception (85), convenience, expected or perceived benefits and adverse effects. Thus, treatments that have been successful in clinical trials may not find the expected success in patient uptake (86).

One reason for this is that, as already mentioned, despite progress in shared medical decision-making, patients have typically had limited opportunities to express their opinions on their disease and treatments in a way that directly affects the pharmaceutical development process. Clinicians and doctors have historically played a gatekeeping role by interfacing directly with patients and translating these interactions into a view of the patients' needs, including clinical rating scales, disease conceptual models and treatment and diagnostic guidelines. This gatekeeping role has influenced the definition and evaluation of what clinical success means, ultimately steering pharmaceutical companies away from a full understanding of the patients' view and towards a clinicians' view of the disease (7, 87). Therefore, it is possible to talk about a "translational gap" between the clinician's view of patients and actual patients treated while being busy with their daily lives.

To increasingly incorporate the patient's view in the pharmaceutical process there has been a growing emphasis in measuring quality of life (QoL) outcomes in clinical trials through patient-reported outcomes (PROs) (88, 89). PROs are patient self-reports that aim at capturing the opinion of patients in ways that can be systematically evaluated and can be used to, among other things, identify changes in disease treatment that lead to the largest improvement in QoL. PROs can present several shortcomings, such as response bias and lack of content validity (90). Moreover, there is a set of specific challenges for PROs in the context of clinical trials due to lack of representation of real-world conditions, reliance on validated and long-established measures that might be outdated, limited patient input in their creation (e.g., with the help of patient organizations or a limited set of patient interviews) and selection bias (as PROs in clinical trials only study the patients who qualify and enroll in the trial and are not a representative patient population sample). Therefore, SML can complement PROs by addressing many of their shortcomings.

4.2 Legal framework challenges

A fundamental legal and ethical problem with open social media data is that it cannot be fully anonymized, because the original source can always be found through an internet search engine. An alternative to anonymization is to create synthetic data that retains the relevant characteristics of the original data, e.g., by paraphrasing, using synonyms or pseudonyms, etc. This can be done manually but requires extensive work. More recently, advances in natural language generation (NLG) with large language models (LLMs) hint at the possibility of creating anonymized synthetic social media data at scale. Until this is a reality, it can be generally assumed that anonymity cannot be accomplished at scale.

The challenge of lack of anonymity plus regulatory and legal uncertainty has deterred pharmaceutical companies from widely adopting SML for PFDD. Thus, while patients have been using the internet since its inception to discuss their health condition, and this usage exploded along with social media channels such as forums, blogs and microblogs (68), only increased clarity in the legal framework together with encouragement from regulatory agencies have fostered the research of SML by pharmaceutical companies (cf. section 2.2). Thus, legal and regulatory clarity have been necessary stepping stones in furthering the adoption of SML in the pharmaceutical industry.

One major challenge has been variability in the laws and their implementation across countries. This requires establishing interpretations and best practices that apply to multiple jurisdictions simultaneously, particularly given the global nature of online content, data hosting and pharmaceutical industry locations. Thus, with the exception of activities circumscribed to a single country (e.g., Chinese-internal SML), it is necessary to follow the principle of “most restrictive rule” among all overlapping legal frameworks, whereby the guidelines to be followed are derived from the set of most restrictive applicable laws.

In the EU, the General Data Protection Regulation (GDPR) established a generally more restrictive legal framework for SML that is being increasingly taken as a reference by legislation in non-EU countries. The GDPR seeks to balance legitimate interests for data use with the fundamental rights and freedoms of individuals, as well as their expectations. The GDPR also introduces certain exemptions for scientific research purposes that support a broader use of SML, including, if scientifically justified, inferring patient information not explicitly stated. To qualify for this research exemption the following conditions would need to be met (91): (1) relevant sectoral standards of methodology and ethics are followed, and (2) the research is carried out with the aim of growing society’s collective knowledge and wellbeing, as opposed to serving primarily private interests. The GDPR includes, moreover, regulations in terms of data storage and processing that also apply to research activities (92, 93).

If SML is performed as a research activity, nonetheless, certain local regulations beyond GDPR may apply that require an approval of the activities by corresponding research ethics committees, which in the US are called institutional review boards (IRBs). IRB approval in the US, however, is not required if it only involves “observation of public information” (94). It has been debated whether open social media sites that require user authentication can still be considered public. In such cases, the expectation of the users and the terms and

conditions of the site may help to determine this. A simpler, but more restrictive, approach is to automatically disregard any site which requires a password (95). In other jurisdictions besides the US, research ethics committees may hold somewhat different policies towards observational research.

Even for research activities, it is generally advisable to process data that is “manifestly made public” to minimize potential harm, as prescribed by the GDPR for the processing of certain types of personal data. This precludes inferring information about individuals that is not openly stated (e.g., inferring that a patient has a disease from the writing style) and integrating data about specific patients coming from different sites. Overall, if SML is not performed as a research activity, the GDPR emphasizes the importance of protecting the fundamental rights of patients.

Another legal aspect to be considered concerns the terms and conditions of each social media site, which determine the range of activities that are allowed with the site’s content. Terms and conditions vary widely across sites and can be unclear about data reuse. Thus, sites can range from early-internet-style communities in which there are no explicit data reuse restrictions, to sites from for-profit companies that offer the reselling of user data, which many of the site users may have failed to notice. Indeed, many users might find it difficult to judge a site’s terms and conditions, whether because these are difficult to locate, or because of their length and complexity. Thus, there is a contrast between the care with which terms and conditions need to be considered by those performing SML and the lack of attention to them paid by many of the users who contribute the content. Reviewing the legal and ethical behavior of social media sites should be considered an integral part of the SML process. Additionally, since terms and conditions can change over time, a best practice is to record them and archive them every time SML activities are being performed.

4.3 Ethical considerations

Beyond grappling with legal aspects, pharmaceutical companies need to address ethical questions regarding SML, which can present different urgency across countries and cultures. While the ethnography of online communities is an established field of academic research (96), users have diverging opinions about its appropriateness and this opinion can be highly context-dependent (97, 98). Thus, when it comes to research with the goal of improving healthcare, opinions can be more supportive (99, 100), as discussed in section 3. Despite that, there can be additional concerns when such research is done by pharmaceutical companies.

Such ethical aspects can be tackled if we consider SML studies within the ethical framework of observational studies. It has been a matter of debate whether observational studies can be ethical without patient consent, which is the bedrock of modern clinical practice (101), and which conditions would need to be met for the absence of patient consent to be ethical. Thus, for example, UK’s Economic and Social Research Council (ESRC) considers ethical the forgoing of patient consent as long as a study is not “undertaken lightly or routinely. It is only justified if important issues are being addressed and if matters of social significance which cannot be uncovered in other ways are likely to be discovered” (102). It could be argued that SML can fulfill those requirements when done appropriately.

Additionally, the ESRC justifies forgoing the requirement of informed consent when “overt observation might alter the phenomenon being studied” (102). Indeed, the feasibility of obtaining informed consent in SML would hinge on the ability of communicating with the patients, or caregivers, who post online, which might be unfeasible. Additionally, the act of contacting users may bias their own future online actions as well as those of their peers, as they may start behaving differently, especially if they have a negative pre-established view of pharmaceutical companies. Therefore, contacting users to notify them of SML activities could change the nature of online conversations in a way that could be detrimental to future SML activities by any researchers, whether pharmaceutical or not.

A broader review of ethical recommendations (103) suggests that SML for health research can be ethical without informed consent when there is an emphasis on ensuring user anonymity, minimization of possible harms to users, a focus on public benefit, transparency in data access and analysis methods, and abidance to the law and terms and conditions from the source sites. Beyond these recommendations, there could be additional ethical questions if the research is performed by for-profit corporations, such as pharmaceutical companies. For example, patients could feel entitled to a share in the profit resulting from sharing their data online. While payments for participation in research studies are considered acceptable to compensate for study burdens and out-of-pocket expenses, or to attract participation (104), there has been less discussion about payments associated with for-profit research. While potentially reasonable, this question could only be addressed if the economic benefits derived from insights from SML could be quantified with some degree of precision.

Ethical discussions about SML have mainly focused on its potential negative impacts on patients, neglecting consideration of the ethics of its positive impacts. Widely regarded principles of medical ethics (105) include both the duty of non-maleficence (“doing no harm”) and the duty of beneficence (“doing good”). Hence, one could argue that pharmaceutical companies that do not try to derive evidence from freely available social media data are doing a disservice to patients. However, it is important to note that there is no moral or legal imperative in medicine to use all existing data, and medical neglect is only applicable when established medical practices are not followed. However, social media data is qualitatively different from other types of patient-related data as it can be considered as representing the “voice of the patient.” For instance, patients may judge the neglect of a genomic dataset that remains unanalyzed differently from the dismissal of complaints about long COVID on social media. One potential solution to this ethical question is maybe not to make SML a mandatory practice but to consider it a best practice, which could fit within the framework of evidence-based medicine, which aims to systematically use the best available evidence.

In fact, patients want their doctors to be omniscient (“all knowing”) (106) and, while omniscience in medicine may appear as an unreasonable demand, it is not completely unachievable. There is already a legal mandate by which pharmaceutical companies have the duty to actively seek all available data in one specific realm: pharmacovigilance. Pharmacovigilance focuses on monitoring potential adverse events of marketed drugs (107, 108) and, in this realm, social media is already acknowledged as a source of relevant patient data that can play a complementary role to existing approaches in certain cases (109, 110). Expanding the analysis of social media data

beyond pharmacovigilance, would be an acknowledgement that other aspects of medical treatments are as important to patients as adverse events.

4.4 Summary

Actively evaluating information from patients about PFDD-relevant topics as expressed in social media channels should be considered a pursuit that can have a strong impact on broadening the portfolio of methods beyond existing PRO frameworks. Moreover, SML can exhibit a positive ethical value when done appropriately. In that regard, from the experience of the authors of this article with conducting SML studies, and in line with current practice from the literature (cf. section 3), we recommend the principles stated in Table 2 as best practices when conducting PFDD-related SML studies in a pharmaceutical industry setting.

5 Considerations by regulatory authorities

Regulatory authorities worldwide have embraced the challenge of adapting their frameworks to ensure that patients’ needs and preferences are systematically taken into account into drug development. The FDA for instance has defined PFDD as a “systematic approach to help ensure that patients’ experiences, cultural traditions, perspectives, needs, concerns and priorities are captured and meaningfully incorporated into drug development and evaluation.” The main goal of this initiative is to contribute to improved health outcomes for different groups of patients.

In the following, we provide an overview of the existing positions of different regulators (FDA, EMA, MHRA) on the systematic involvement of patients in drug development to the extent that they are publicly available (section 5.1). Subsequently, we specifically discuss their views on how SML approaches can support PFDD initiatives and how they could potentially complement existing

TABLE 2 Best practices for the conduct of PFDD-related SML studies by pharmaceutical companies.

- Follow the principle of “most restrictive rule” among all applicable legal frameworks.
- Prioritize user anonymity and minimize any potential negative effect on users.
- Consider the possibility of creating synthetic data that reflects the characteristics of the original data but maintains anonymity.
- If possible, do not integrate patient data across social media sites.
- If possible, do not infer patient data, only use data openly stated.
- Disregard sites which require login and password.
- Regularly review and adhere to sites’ terms and conditions.
- Systematically record and archive the terms and conditions of each site for every SML activity.
- Evaluate the ethical and legal behavior of sites towards their users.
- Ensure secure data storage practices to safeguard collected information.
- Control access to stored social media data, limiting it to authorized personnel with a legitimate need for analysis.
- Regularly audit and assess data storage and access protocols to maintain compliance with evolving security and legal standards.

regulatory decision making practices (section 5.2). We finally raise important questions that need to be clarified regarding the role and impact that social media can have on regulatory decision making.

5.1 Patient-focussed drug development initiatives

5.1.1 Food and Drug Administration

The FDA has so far played a pioneering role in the PFDD initiative, being the first regulatory body having published guidelines on how patient experience can be incorporated into drug development. As a consequence of the 21st Century Cures Act of 2016 and the FDA Reauthorization Act of 2017, the agency has issued a series of guidelines for industry and stakeholders providing recommendations on how patient experience data could be incorporated into PFDD:

- Collecting Comprehensive and Representative Input [(111), Final Guidance available as of June 2020].
- Methods to Identify What Is Important to Patients [(112), Final Guidance available as of February 2022].
- Selecting, Developing or Modifying Fit-for-Purpose Clinical Outcomes Assessments [(113), Draft Guidance available since June 2022].
- Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision Making [(114), Draft Guidance available since April 2023].

In accordance with the Cures Act, the FDA has conducted assessments of its use of patient experience data in 2021 and follow-up assessments are planned for 2026 and 2031. The 1st assessment has been carried out in cooperation with ERG and documented correspondingly (81). The assessment has emphasized that there is still significant uncertainty regarding how exactly the FDA might use patient experience data and it is frequently emphasized that the FDA and industry are currently “*in the middle of a learning curve*” ahead of understanding and developing best practices. When the FDA uses patient experience data in regulatory decision making, it usually takes the form of considering PROs and other clinical outcome assessments (COAs) as primary endpoints in the benefit-risk analysis for a marketing application as well as background context for the review.

For drug and biologic marketing applications received after June 12, 2017, the FDA has been fulfilling the requirement to make a public statement about its use of patient experience data by including a Patient Experience Data Table in review documents for approved applications. As of June 2017, the FDA is required to publish a brief statement about how patient experience data that was part of a drug biologic application was used as part of the review.⁴ From all NME NDA/BLA reviews that mention patient experience data, at the time of writing, 88% contain such a Patient Experience Data Table.

When using patient experience data for PFDD, the following aspects have been identified as best practice (81):

- Early and frequent communication between FDA and applicants.
- Development of a solid data analysis plan.
- Applicant use of patient experience data to help design clinical trials.
- FDA to use patient experience data to frame the review.
- Sharing of patient experience data with the patient community.

It has been stated in general that applicants need more clarity and guidance on how patient experience data can be used as background or context, as well as for benefit-risk analysis.

5.1.2 European Medicines Agency

In their strategic reflection paper “EMA Regulatory Science to 2025,” the EMA has generally stressed the importance of patient involvement in drug development (13). As part of the goal 2 on “*Driving collaborative evidence generation—improving the scientific quality of evaluations*,” the importance of including patient preferences to inform benefit-risk assessment has been emphasized, as well as the identification of areas of high unmet need. The paper emphasizes in general that “*patient perspectives are particularly important, and their involvement can greatly improve trial design and conduct, and the usefulness of the results and medicines developed*.” A goal is thus to foster the input of patients/patient representatives and carers in the regulatory process. As part of goal 3 on “*Advancing patient-centered access to medicines in partnership with healthcare systems*,” the EMA has stressed that it is looking to enhance its methodological portfolio to enable greater input from the wider patient community in a systematic manner. It has mentioned in particular the aim to “*explore and deploy additional methodologies to collect and use of patient data for benefit-risk assessment*.” It has highlighted in particular that new digital technologies have the potential of providing a more holistic view of the patient and the disease and that “*If analyzed appropriately, these new sources of data can create new evidence which has the potential to add significantly to the way the benefit-risk of medicinal product is assessed over their entire lifecycle*.”

In a recent joint reflection paper of the EMA and the International Committee for Harmonisation (ICH), the agencies have stressed that “*it is increasingly critical to develop a harmonized approach to collecting and incorporating patient perspectives for these to become more prominent in drug development and decision making*” (1). The authors have emphasized that to maximize benefit of patient perspectives in these areas, regulators and drug sponsors need to employ methods and measures that:

- Include patients and caregivers as partners to best inform the work.
- Ensure the information collected is sufficiently reliable, valid and representative to be used as basis for planning and decision making.
- Can be deployed in a timely and sustainable way.
- Will be relevant to patients and their caregivers.
- Account for heterogeneity of groups.

In line with the FDA, the EMA/ICH has highlighted the importance of identifying important impacts and concepts from patients as a basis to select, modify or develop COAs that can demonstrate meaningful change in patients’ lives. The EMA/ICH

⁴ This may include using patient experience data for the purposes of supporting interpretation or summary of the data, enhancing data analysis, or as a decisive factor in bringing the final decision about.

specifically lists the questions that could be answered from the patients' perspective:

- What disease burdens and treatment effects matter most to patients that might be addressed by a medical theory?
- What would be the best way to measure these disease or treatment burdens/effects in a clinical trial, and are the methods acceptable for patients?
- What would be the most appropriate endpoints to use in clinical trials (and robust enough to inform regulatory decision making)?
- What are clinically meaningful changes in an endpoint from a patient perspective?

As an interim conclusion, it can be stated that the main use case for patient experience data is to inform the definition, development and validation of PROs and other types of COAs. It has been particularly noted that patient experience data can play a role in “*providing supporting information in situations where the condition is not well characterized, as with some rare diseases*” (81).

5.1.3 Medicines and Healthcare products Regulatory Agency

The MHRA has also emphasized the importance of the inclusion of patient perspectives in drug development in their strategy paper on “*Saving and Improving Lives: The Future of UK Clinical Research Discovery*” (14). One of the five pillars of the MHRA vision for UK clinical research delivery is patient-centered research. The MHRA has acknowledged that “*patients must be routinely involved in the design of clinical research to ensure outcomes match their needs and studies are designed with real participants and the realities of their daily lives in mind.*” As a first step in this direction, the MHRA has started a pilot in March 2021 to support the strategic goal of integrating the perspective of patients into the decision making process regarding the approval of new medicines. As part of this pilot, applicants will be specifically requested to provide evidence on the patient involvement activities undertaken when developing their product. The main goal for MHRA is to learn from the evidence submitted to define their strategy.

5.2 Regulatory perspectives on social media listening in patient-focused drug development

In light of the observable convergence of regulatory authorities in giving high priority to enforcing the involvement of patients into drug development and decision making processes (as summarized in the previous subsection), it is a natural question to ask to what extent SML approaches may qualify, from the regulators' perspective, to elicit the required patient experience data from patients directly to make PFDD a reality. Given that the most articulate position in this regard has been put forward by the FDA to date, we focus our subsequent discussion on their perspective.

In fact, social media research has been identified as a potential method for generating experience data in the various guideline documents on PFDD issued by FDA. In their guidelines, FDA has recommended that patient experiences are captured as directly

reported by patients rather than through mediation or interpretation of others.

SML is considered to satisfy the requirement of capturing the direct input of patients, and in this sense is comparable to interviews carried out directly with the patient (1-on-1, deep or cognitive interviews, concept elicitation interviews etc.). SML captures the spontaneous and unsolicited input of patients as shared by users while being active on social media. As discussed in section 2, SML does not suffer from common biases inherent in the design of a questionnaire (leading questions, bias due to order of questions etc.).

While 1-on-1 interviews and focus group studies are carried out with between 5 and 20 participants, SML has the advantage of capturing the perspective of a more diverse and broad patient population. It can thus contribute to the representativeness of patient experience data across the full diversity of the patient population and would help to fulfill recent draft FDA guidance requirement to provide “*Diversity Plans to Improve Enrollment of Participants from Underrepresented Racial and Ethnic Populations in Clinical Trials*” (115).

The benefits of SML recognized by the FDA are the following:

- May allow access to hard-to-reach populations.
- Cost and time efficient.
- Easy to implement.
- Accurate and automatic capture of data.
- Low burden on participants.

In general, SML should not stand on its own, but complement existing methods such as one-on-one interviews or focus group studies or other survey data. The FDA has mentioned mixed method designs [cf. (116)] that allow different research instruments to be applied in a synergistic manner, thus complementing each other. We spell out below how social media patient listening fits into the envisioned synergetic use of different methods identified by the FDA:

- Harmonizing and confirming results from different methods.
- Supplementing and clarifying results from one method with results from another method.
- Using results from one method to inform the design of another method.
- Discovering inconsistencies, contradictions and new perspectives, and reframing of questions or results from one method with questions or results from the other method.
- Expanding the scope of research questions by using different methods for different components of the research question.

Overall, the FDA has so far refrained from giving precise guidance on how exactly SML should be used or combined with different methods for the purposes of fostering PFDD. This holds for methodological aspects related to data selection and analysis, and more technical or privacy-related issues such as data anonymization or redaction of personally identifiable information as well. On the other hand, the agency consistently emphasizes that all methods and data put forward for regulatory decision making—with patient experience data originating from SML being explicitly included (112)—will always be assessed from the perspective of whether they are fit for purpose to address and answer the research question at

stake. In all these regards, the FDA does not distinguish between automated or manual approaches to SML.

5.3 Summary

Summarizing the regulatory considerations, it is fair to say that regulators acknowledge the need for new methods and techniques to reliably capture the experiences of patients and what is important to them. The FDA in particular has highlighted that social media could complement more established qualitative and quantitative research methods by reliably capturing patient insights while minimizing their burden.

Key methodological challenges, as mentioned in section 2, are inaccuracies, reliance on patients' self-disclosure, missing data, privacy protection. The main benefits seen by regulators are: (i) enhanced access to hard-to-reach populations, (ii) cost and time efficiency, (iii) ease of implementation, (iv) low burden for participants, and (v) access to the direct and unmediated voice of the patient. From a methodological perspective, ensuring the generalizability to the target population while minimizing biases to particular segments of the population is a key challenge seen by regulatory bodies.

Understanding how exactly social media patient listening can complement existing research instruments beyond the current general statements made by FDA is an important priority. As directions for future research, we envisage the following potential use cases that need to be spelled out in more detail together with their specific methodological underpinnings in order to provide clearer guidance on how social media derived guidance can impact regulatory decision making:

- Social media as background on the impact of the disease on patients (natural history of disease).
- Social media to establish an unmet need from the patients' perspective.
- Social media as a way to define and demonstrate the relevance of a clinical endpoint from the patients' real-world perspective.
- Social media as a way to inform the assessment of the benefit-risk tradeoff of a medication.

6 Synthesis

Patients are often concerned that research and development for new medicines is disconnected from their needs and priorities. Indeed, there is an increasing recognition that involving patients in research and development of medicines can provide significant value to all stakeholders involved. Patients have a unique perspective on what it means to live with a particular condition, and this perspective can be invaluable in drug development. Patients can provide insights into the impact of a particular condition on their daily lives, including symptoms, quality of life, and the practical challenges they face, and may also help to design, improve recruitment and retention in clinical trials.

An important question, certainly, is how to include the perspectives of patients in a manner that is effective, efficient, reduces

the burden of patients and ensures that their perspectives are accurately taken into account, while the approach for doing so is legitimate and faithfully captures the perspectives of patients. There are two important considerations here. For one, Haerry et al. (3) have discussed that what is important is to include the naive and "lay perspective" of patients in the process as the beneficiaries of drug development, as they are the only ones that from their lay perspective understand which improvements are needed, without necessarily having to understand the methodological, scientific, regulatory and practical aspects involved in bringing these improvements about. It is exactly this "lay perspective" that is most relevant for drug development. Haerry et al. (3) have highlighted a paradox in that the more patients are involved in the process and the more educated they get as part of this engagement, the less they represent a naive and lay perspective. For another, it is key to minimize the burden of patients associated with their involvement in drug development, which is considerable in the case of relying on PRO instruments (117). Thus, patients can bring both an expert understanding related to their disease and patient journey as well as lay perspective by not being healthcare professionals.

Social media serves an important function for patients providing informational, emotional and social support and a conducive environment for patients to disclose sensitive and private experiences through anonymity. Within this anonymity, they are willing to share important information about their disease, how it impacts their lives, the challenges they face and which unmet needs still exist. Thus, listening to patients on social media offers an immense opportunity to gain access to the authentic, unfiltered, unbiased and first-person experience of patients in addition to providing a more holistic view of patients by highlighting their social, cultural and emotional contexts. It is thus providing us access to what Haerry et al. (3) have called the "lay perspective" of patients, while minimizing the burden for them to be involved in drug development.

A thorough understanding of patient experiences and priorities is key to narrowing down the gap that exists between the treatments on the market and under development, and the actual needs of patients and the outcomes that would have a significant positive impact on their lives.

Trial sponsors are clearly embracing the challenge of developing treatments that are relevant to patients and have recently been active in exploring the potential of social media to learn about patients' needs, as this article, and the systematic review carried out within it, clearly corroborate. Methodologies for analyzing the perspective of patients from social media are maturing, and a clear convergence in terms of data protection practices can be observed. Best practices exist that allow to carry out patient SML studies in a way that fosters anonymity to protect patients' privacy rights and relies on content that has been made manifestly public. SML can be carried out in a way that is ethical even without explicit consent if it is done for research purposes, not conducted routinely, and if the insights have a positive societal impact and cannot be gained or uncovered easily in other ways. Developing new treatments that alleviate the burdens of patients is without doubt of high societal interest, so that one could argue that there could be an ethical obligation to listen to patients' voices on social media.

Regulatory authorities have consistently highlighted the importance of inclusion of patient perspectives in drug development in order to maximize real-world outcomes that positively impact

patients' lives. While EMA and MHRA are clearly committed and even mandated the inclusion of patient perspectives in drug development, the FDA has provided a set of regulatory guidance documents to specify their methodological expectations on how patients should be included in drug development. The EMA has clearly stressed that “new technologies” should be explored and the FDA has explicitly included SML as one of the important methods to capture patient experience data. Yet, the FDA has also stressed that as far as the use of social media is concerned, we are all part of a learning curve. The main issue at stake is to understand how representative the perspectives are and how generalizable they are to the actual target populations. In general, we need to understand and balance the methodological tradeoffs between the benefits and drawbacks of relying on social media. At one end of the spectrum, online populations might be biased to certain segments of the patient population, data collection might be affected by selection bias and might be incomplete as patients would not spontaneously provide information on all aspects relevant for a certain question. Furthermore, information on social media cannot be verified. In this sense, the reliability of social media is not comparable to data rigorously collected in HRQoL studies or PRO studies, but it is comparable to other sources of RWE and could certainly complement or inform the design of such instruments and make sure that important aspects are captured, contributing to content validity. Additionally, SML has the potential to access large numbers of patients, even for rare conditions, can help to get the real, undistorted voice of patients, can ensure that we capture what is relevant from their perspective and minimizes the burden of collecting patient experience data. Future activities should be devoted to developing methodological approaches and best practices that balance the above mentioned drawbacks and benefits. Regulators have stressed that their ability to provide regulatory guidance is limited and that they will accept any method that is fit for purpose. It is thus in the hands and responsibility of the community of sponsors, tech vendors and patient organizations to define methods that are fit for purpose in dialog with the regulators.

Overall, the interests of the three stakeholder groups that we have discussed in this article are clearly aligned. Patients want to have a stake in drug development and their voices to be considered from early on. Social media is a way to do so while minimizing their burden. Sponsors have already invested in developing methodologies to capture the online patient experience as it gives them access to the first hand and direct experience of patients and is a cost effective and scalable method to incorporate their perspective into drug development. Regulatory bodies are explicitly acknowledging social media as an important source of patient experience data. What is missing is a regulatory framework and policies that create certainty on how patient experience data collected from social media can be used in ways that are aligned and compliant with the interest and requirements of the three groups discussed here.

7 Recommendations

Patients, regulatory bodies and industry need to work collaboratively to accelerate PFDD. From the perspective of all stakeholders, legitimacy, privacy and compliance as well as methodological robustness seem to be important factors. These open points define a clear roadmap that the above stakeholders should work

on together to reduce regulatory, legal and methodological uncertainty in ensuring that perspectives of patients are captured in drug development from early on to reduce the gap between what is measured in current clinical trials and what would really make a difference to the daily living and functioning of patients.

In order to leverage SML methods even further as a source of real-world patient experience data to support patient-focused drug development, more concise regulatory provisions, best practices and guidelines are needed that reduce uncertainty for all stakeholders on the following issues in particular:

1. Data collection: Regulatory recommendations and guidelines are needed on how to collect data from social media in a way that is compliant with established provisions on privacy and data protection (GDPR in particular) and current ethical standards. This also includes recommendations on how to collect data to minimize selection biases.
2. Data analysis: Standardized and robust methods are needed to generate results that generalize to the target population including measures to convincingly demonstrate robustness of analyzes and highlighting potential selection biases transparently. This includes recommendations on what level of evidence is required to assume, e.g., that a patient with a self-disclosed diagnosis fulfills the inclusion criteria, or that a particular variable of interest can be reliably attested based on patients' self-reports.
3. In addition to the rationale from industry and regulators, justification from patients or patient advocacy groups must be elicited for the legitimacy and validity of using SML to capture PFDD-related insights.
4. Accepted methods for using social media data to inform the design of PRO instruments, HRQoL surveys, interviews or other methods to elicit meaningful patient input for PFDD.
5. Guidelines and best practices on how results from social media analysis can enhance or support decision making on which endpoints to include in a clinical trial.
6. Guidelines and best practices on how results from social media analysis can support benefit-risk analysis as part of regulatory approval.

As part of their agenda, the Expert Community Group “Exploiting Real-World Data from Social Media in Patient-Focused Drug Development” within the Pistoia Alliance intends to work on these topics to contribute to methodological clarification and to the development of best practices.

Author contributions

PC: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. BC: Writing – review & editing. MDV: Investigation, Writing – review & editing. TE: Funding acquisition, Writing – review & editing. JG: Writing – review & editing. MH: Investigation, Methodology, Writing – original draft, Writing – review & editing. ML: Conceptualization, Writing – review & editing, Writing – original draft. BN: Writing – review & editing. RR: Writing – original draft. ALS: Writing – review & editing. CSK: Investigation, Methodology, Writing – original draft, Writing – review & editing. MV: Writing – review & editing. KW: Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

This work has been carried out by the Expert Community Group “Towards Regulatory Standards for Exploiting Real-World Data from Social Media in Patient-Focused Drug Development Community” of the Pistoia Alliance. The authors thank all group members for comments and input on the manuscript and discussions.

Conflict of interest

PC, MH, and CS-K were employed by Semalytix GmbH. BC and MVo were employed by Boehringer Ingelheim GmbH. MVu was employed by Chiesi Farmaceutici SpA. JG, ML, RR-E, and AS were employed by F. Hoffmann-La Roche Ltd. BN was employed by Takeda

References

- ICH (2021). Proposed ICH guideline work to advance patient focused drug development. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-reflection-paper-proposed-ich-guideline-work-advance-patient-focused-drug-development-pfdd_en.pdf. (Accessed June 1, 2023)
- Lowe MM, Blaser DA, Cone L, Arcona S, Ko J, Sasane R, et al. Increasing patient involvement in drug development. *Value Health*. (2016) 19:869–78. doi: 10.1016/j.jval.2016.04.009
- Haerry D, Landgraf C, Warner K, Hunter A, Klingmann I, May M, et al. EUPATI and patients in medicines research and development: guidance for patient involvement in regulatory processes. *Front Med*. (2018) 5:230. doi: 10.3389/fmed.2018.00230
- Anderson M, McCleary KK. From passengers to co-pilots: patient roles expand. *Sci Transl Med*. (2015) 7:291fs25. doi: 10.1126/scitranslmed.aac6023
- Bartlett SJ, Barnes T, McIvor RA. Integrating patients into meaningful real-world research. *Ann Am Thorac Soc*. (2014) 11:S112–7. doi: 10.1513/AnnalsATS.201309-327RM
- Boutin M, Dewulf L, Hoos A, Geissler J, Todaro V, Schneider RF, et al. Culture and process change as a priority for patient engagement in medicines development. *Ther Innov Regul Sci*. (2017) 51:29–38. doi: 10.1177/2168479016659104
- Morel T, Cano SJ. Measuring what matters to rare disease patients—reflections on the work by the IRDiRC taskforce on patient-centered outcome measures. *Orphanet J Rare Dis*. (2017) 12:171. doi: 10.1186/s13023-017-0718-x
- Carson RT. Patient experience data for medical product development: opportunity beyond obligation. *Patient*. (2022) 15:147–9. doi: 10.1007/s40271-021-00570-8
- Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. *Trials*. (2017) 18:122. doi: 10.1186/s13063-017-1870-2
- Molnar FJ, Man-Son-Hing M, Fergusson D. Systematic review of measures of clinical significance employed in randomized controlled trials of drugs for dementia: clinical significance in dementia drug research. *J Am Geriatr Soc*. (2009) 57:536–46. doi: 10.1111/j.1532-5415.2008.02122.x
- Rahimi K, Malhotra A, Banning AP, Jenkinson C. Outcome selection and role of patient reported outcomes in contemporary cardiovascular trials: systematic review. *BMJ*. (2010) 341:c5707. doi: 10.1136/bmj.c5707
- Mullin TM (2013). FDA's patient-focused drug development initiative. Available at: https://c-path.org/wp-content/uploads/2013/09/PRO_Consortium_FDA_PDUFA_V_PatientFocusedDD.pdf. (Accessed June 1, 2023).
- EMA (2020). EMA regulatory science to 2025. Available at: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf. (Accessed June 1, 2023).
- MHRA (2021). Patient involvement strategy 2021-05. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1022370/Patient_involvement_strategy.pdf. (Accessed June 1, 2023).

Pharmaceuticals. KW was employed by SageFiber and he is a consultant to Boehringer Ingelheim International GmbH.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2024.1274688/full#supplementary-material>

- Schmidt AL, Rodriguez-Esteban R, Gottowik J, Leddin M. Applications of quantitative social media listening to patient-centric drug development. *Drug Discov Today*. (2022) 27:1523–30. doi: 10.1016/j.drudis.2022.01.015
- Pope C, Ziebland S, Mays N. Analysis In: C Pope and N Mays, editors. *Qualitative research in health care*. Oxford, UK: Wiley Blackwell (2020). 111–33.
- Spies E, Andreu T, Koelling J, Hartung M, Kamudoni P, Park J. PCR251 retrospective social listening study of patients living with systemic lupus erythematosus (SLE): understanding the patient experience. *Value Health*. (2022) 25:S438. doi: 10.1016/j.jval.2022.09.2184
- Spies E, Andreu T, Koelling J, Hartung M, Kamudoni P, Park J. PCR18 an exploratory retrospective social listening study to identify patient experiences associated with cutaneous lupus erythematosus (CLE). *Value Health*. (2022) 25:S393. doi: 10.1016/10.1016/j.jval.2022.09.1953
- Staunton H, Kelly K, Newton L, Leddin M, Rodriguez-Esteban R, Chaudhuri KR, et al. A patient-centered conceptual model of symptoms and their impact in early Parkinson's disease: a qualitative study. *J Parkinsons Dis*. (2022) 12:137–51. doi: 10.3233/JPD-202457
- Delestre-Levai I, Aliberti S, Almagro M, Carnini C, Chalmers JD, George SC, et al. Patients' perspectives on bronchiectasis: findings from a social media listening study. *ERJ Open Res*. (2021) 7:00096–2021. doi: 10.1183/23120541.00096-2021
- Freeman TCB, Rodriguez-Esteban R, Gottowik J, Yang X, Erpenbeck VJ, Leddin M. A neural network approach for understanding patient experiences of chronic obstructive pulmonary disease (COPD): retrospective, cross-sectional study of social media content. *JMIR Med Inform*. (2021) 9:e26272. doi: 10.2196/26272
- Gries KS, Fastenau J. Using a digital patient powered research network to identify outcomes of importance to patients with multiple myeloma. *J Patient Rep Outcomes*. (2020) 4:74. doi: 10.1186/s41687-020-00242-5
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. (2021) 372:n71. doi: 10.1136/bmj.n71
- Chauhan J, Aasaitambi S, Márquez-Rodas I, Formisano L, Papa S, Meyer N, et al. Understanding the lived experiences of patients with melanoma: real-world evidence generated through a European social media listening analysis. *JMIR Cancer*. (2022) 8:e35930. doi: 10.2196/35930
- Faust G, Booth A, Merinopoulou E, Halhol S, Tosar H, Nawaz A, et al. The experiences of patients with adjuvant and metastatic melanoma using disease-specific social media communities in the advent of novel therapies (excite project): social media listening study. *JMIR Cancer*. (2022) 8:e34073. doi: 10.2196/34073
- Tadmouri A, Alivon M, Andreu T, Hartung M, Ryll B, Rauch G, et al. RWD141 exploration of melanoma patient-generated real-world data using an Ai-based social listening approach. *Value Health*. (2022) 25:S476–7. doi: 10.1016/j.jval.2022.09.2366

27. McDonald L, Behl V, Sundar V, Mehmud F, Malcolm B, Ramagopalan S. Validity of social media for assessing treatment patterns in oncology patients: a case study in melanoma. *JAMIA Open*. (2019) 2:416–22. doi: 10.1093/jamiaopen/ooz013
28. Mazza M, Piperis M, Aasaithambi S, Chauhan J, Sagkriotis A, Vieira C. Social media listening to understand the lived experience of individuals in Europe with metastatic breast cancer: a systematic search and content analysis study. *Front Oncol*. (2022) 12:863641. doi: 10.3389/fonc.2022.863641
29. Shimbkhada R, Attai D, Scheitler AJ, Babey S, Glenn B, Ponce N. Using a Twitter chat to rapidly identify barriers and policy solutions for metastatic breast cancer care: qualitative study. *JMIR Public Health Surveill*. (2021) 7:e23178. doi: 10.2196/23178
30. Lognos B, Carbonnel F, Boulze Launay I, Bringay S, Guerdoux-Ninot E, Mollevi C, et al. Complementary and alternative medicine in patients with breast cancer: exploratory study of social network forum data. *JMIR Cancer*. (2019) 5:e12536. doi: 10.2196/12536
31. Ping Q, Yang CC, Marshall SA, Avis NE, Ip EH. Breast cancer symptom clusters derived from social media and research study data using improved K-medoid clustering. *IEEE Trans Comput Soc Syst*. (2016) 3:63–74. doi: 10.1109/TCSS.2016.2615850
32. Hasan A, Levene M, Weston D, Fromson R, Koslover N, Levene T. Monitoring COVID-19 on social media: development of an end-to-end natural language processing pipeline using a novel triage and diagnosis approach. *J Med Internet Res*. (2022) 24:e30397. doi: 10.2196/30397
33. He Q, Du F, Simonsen LWL. A patient journey map to improve the home isolation experience of persons with mild COVID-19: design research for service touchpoints of artificial intelligence in eHealth. *JMIR Med Inform*. (2021) 9:e23238. doi: 10.2196/23238
34. Miyake E, Martin S. Long COVID: online patient narratives, public health communication and vaccine hesitancy. *Digit Health*. (2021) 7:205520762110596. doi: 10.1177/20552076211059649
35. Picone M, Inoue S, DeFelicis C, Naujokas MF, Sinrod J, Cruz VA, et al. Social listening as a rapid approach to collecting and analyzing COVID-19 symptoms and disease natural histories reported by large numbers of individuals. *Popul Health Manag*. (2020) 23:350–60. doi: 10.1089/pop.2020.0189
36. Dobson R, Pasquarelli N, Gafarova M, Eighteen C, Joschko N, Mandel C, et al. The patient perspective on family planning needs and priorities in multiple sclerosis: a combined quantitative and qualitative research study. *Mult Scler*. (2022) 28:130–691. doi: 10.1177/13524585221123687
37. Lan G, Bhat SP, Jaideep S, Gupta T, Kalyanaraman S, Ghanapathy SV, et al. Understanding symptoms and their impact on patients with multiple sclerosis: what we can learn from social media. *Mult Scler*. (2021) 27:134–740. doi: 10.1177/13524585211044667
38. Donkers SJ, Kirkland MC, Charabati M, Osborne LC. Perspectives of people with multiple sclerosis about helminth immunotherapy. *Int J MS Care*. (2020) 22:43–51. doi: 10.7224/1537-2073.2019-044
39. Rezaallah B, Lewis DJ, Pierce C, Zeilhofer H-F, Berg B-I. Social media surveillance of multiple sclerosis medications used during pregnancy and breastfeeding: content analysis. *J Med Internet Res*. (2019) 21:e13003. doi: 10.2196/13003
40. Tyagi N, Aasaithambi S, Chauhan J, George A, Zaour N. PUK33 patient insights for complement 3 glomerulopathy (C3G) using social media listening. *Value Health*. (2019) 22:S919. doi: 10.1016/j.jval.2019.09.2717
41. Goodspeed K, Mosca LR, Weitzel NC, Horning K, Simon EW, Pfaller AC, et al. A draft conceptual model of SLC6A1 neurodevelopmental disorder. *Front Neurosci*. (2023) 16:1026065. doi: 10.3389/fnins.2022.1026065
42. Adobes-Martin M, Montoya-Morcillo M-L, Zhou-Wu A, Garcovich D. Invisalign treatment from the patient perspective: a Twitter content analyzes. *J Clin Exp Dent*. (2021) 13:e376–82. doi: 10.4317/jced.57835
43. Cook N, Kostikas K, Gruenberger J-B, Shah B, Pathak P, Kaur VP, et al. Patients' perspectives on COPD: findings from a social media listening study. *ERJ Open Res*. (2019) 5:00128–2018. doi: 10.1183/23120541.00128-2018
44. Rodrigues A, Chauhan J, Sagkriotis A, Aasaithambi S, Montrone M. Understanding the lived experience of lung cancer: a European social media listening study. *BMC Cancer*. (2022) 22:475. doi: 10.1186/s12885-022-09505-4
45. Emmott R, Barber SK, Thompson W. Antibiotics and toothache: a social media review. *Int J Pharm Pract*. (2021) 29:210–7. doi: 10.1093/ijpp/riaa024
46. Cotté F-E, Voillot P, Bennett B, Falissard B, Tzourio C, Foulquié P, et al. Exploring the health-related quality of life of patients treated with immune checkpoint inhibitors: social media study. *J Med Internet Res*. (2020) 22:e19694. doi: 10.2196/19694
47. Crawford R, Gries KS, Morrison R, Valluri S, Doward L, Fastenau J. The patient experience of relapsed refractory multiple myeloma and its treatment: a social media review. *Blood*. (2020) 136:16–6. doi: 10.1182/blood-2020-142489
48. Dave A, Yi J, Boothe A, Brashear H, Byrne J, Gad Y. Listening to the HysterSisters: a retrospective keyword frequency analysis of conversations about hysterectomy recovery. *JMIR Perioper Med*. (2019) 2:e10728. doi: 10.2196/10728
49. Perella C, Steenackers M, Robbins B, Stone L, Gervais R, Schmidt T, et al. Patient experience of Sjögren's disease and its multifaceted impact on patients' lives. *Rheumatol Ther*. (2023) 10:601–14. doi: 10.1007/s40744-023-00531-7
50. Flurie M., Coe J., Converse M., Davidson K., Flowers C., Gavigan K., et al. (2022). Real-world evidence from social media provides insights into patient mental health outcomes in the management of gout (abstract) *Arthritis Rheumatol* 74. Available at: <https://acrabstracts.org/abstract/real-world-evidence-from-social-media-provides-insights-into-patient-mental-health-outcomes-in-the-management-of-gout/>. (Accessed June 1, 2023)
51. Silverberg JI, Feldman SR, Smith Begolka W, Johnson MB, Stanojev C, DiBonaventura M, et al. Patient perspectives of atopic dermatitis: comparative analysis of terminology in social media and scientific literature, identified by a systematic literature review. *Acad Dermatol Venereol*. (2022) 36:1980–90. doi: 10.1111/jdv.18442
52. James G, Nyman E, Fitz-Randolph M, Niklasson A, Hedman K, Hedberg J, et al. Characteristics, symptom severity, and experiences of patients reporting chronic kidney disease in the PatientsLikeMe online health community: retrospective and qualitative study. *J Med Internet Res*. (2020) 22:e18548. doi: 10.2196/18548
53. Humphrey L, Willgoss T, Trigg A, Meysner S, Kane M, Dickinson S, et al. A comparison of three methods to generate a conceptual understanding of a disease based on the patients' perspective. *J Patient Rep Outcomes*. (2017) 1:9. doi: 10.1186/s41687-017-0013-6
54. Tapi Nzali MD, Bringay S, Lavergne C, Mollevi C, Opitz T. What patients can tell us: topic analysis for social media on breast cancer. *JMIR Med Inform*. (2017) 5:e23. doi: 10.2196/medinform.7779
55. Reuter K, Lee D. Perspectives toward seeking treatment among patients with psoriasis: protocol for a Twitter content analysis. *JMIR Res Protoc*. (2021) 10:e13731. doi: 10.2196/13731
56. Booth A, Bell T, Halhol S, Pan S, Welch V, Merinopoulou E, et al. Using social media to uncover treatment experiences and decisions in patients with acute myeloid leukemia or myelodysplastic syndrome who are ineligible for intensive chemotherapy: patient-centric qualitative data analysis. *J Med Internet Res*. (2019) 21:e14285. doi: 10.2196/14285
57. Cook N, Mullins A, Gautam R, Medi S, Prince C, Tyagi N, et al. Evaluating patient experiences in dry eye disease through social media listening research. *Ophthalmol Ther*. (2019) 8:407–20. doi: 10.1007/s40123-019-0188-4
58. Webb Hooper M, Mitchell C, Marshall VJ, Cheatham C, Austin K, Sanders K, et al. Understanding multilevel factors related to urban community trust in healthcare and research. *Int J Environ Res Public Health*. (2019) 16:3280. doi: 10.3390/ijerph16183280
59. Teague SJ, Shatte ABR, Weller E, Fuller-Tyskiewicz M, Hutchinson DM. Methods and applications of social media monitoring of mental health during disasters: scoping review. *JMIR Ment Health*. (2022) 9:e33058. doi: 10.2196/33058
60. Hsuen Y, Naslund JA, Brownstein JS, Hawkins JB. Monitoring online discussions about suicide among Twitter users with schizophrenia: exploratory study. *JMIR Ment Health*. (2018) 5:e11483. doi: 10.2196/11483
61. Muchnik L, Pei S, Parra LC, Reis SDS, Andrade JS Jr, Havlin S, et al. Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Sci Rep*. (2013) 3:1783. doi: 10.1038/srep01783
62. Han JY, Kim J-H, Yoon HJ, Shim M, McTavish FM, Gustafson DH. Social and psychological determinants of engagement with an online breast cancer support group: posters, lurkers, and nonusers. *J Health Commun*. (2012) 17:356–71. doi: 10.1080/10810730.2011.585696
63. Farrar M, Lundt L, Franey E, Yonan C. Patient perspective of tardive dyskinesia: results from a social media listening study. *BMC Psychiatry*. (2021) 21:94. doi: 10.1186/s12888-021-03074-9
64. McDonald L, Malcolm B, Ramagopalan S, Syrad H. Real-world data and the patient perspective: the PROMISE of social media? *BMC Med*. (2019) 17:11. doi: 10.1186/s12916-018-1247-8
65. Patalano F, Gutzwiller FS, Shah B, Kumari C, Cook NS. Gathering structured patient insight to drive the PRO strategy in COPD: patient-centric drug development from theory to practice. *Adv Ther*. (2020) 37:17–26. doi: 10.1007/s12325-019-01134-x
66. Rocha HM, Savatt JM, Riggs ER, Wagner JK, Faucett WA, Martin CL. Incorporating social media into your support tool box: points to consider from genetics-based communities. *J Genet Counsel*. (2018) 27:470–80. doi: 10.1007/s10897-017-0170-z
67. Ugwuodike P, Sánchez-Benitez Y. Critical social media analysis: problematising online policy representations of the impact of imprisonment on families. *Int J Offender Ther Comp Criminol*. (2022) 68:0306624X2210865:235–56. doi: 10.1177/0306624X221086559
68. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *J Med Internet Res*. (2013) 15:e85. doi: 10.2196/jmir.1933
69. Bunting AM, Krawczyk N, Lippincott T, Gu Y, Arya S, Nagappala S, et al. Trends in fentanyl content on reddit substance use forums, 2013–2021. *J Gen Intern Med*. (2023) 38:3283–7. doi: 10.1007/s11606-023-08256-7
70. Chochol MD, Gandhi K, Elmaghraby R, Croarkin PE. Harnessing youth's engagement with mental health TikTok and its potential as a public health tool. *J Am Acad Child Adolesc Psychiatry*. (2023) 62:710–2. doi: 10.1016/j.jaac.2022.11.015

71. PWC Health Research Institute (2012). Social media “likes” healthcare: from marketing to social business. Available at: <https://www.pwc.com/us/en/health-industries/health-research-institute/publications/pdf/health-care-social-media-report.pdf>. (Accessed June 1, 2023).
72. Silver L, Huang Christine, Taylor K. (2023). In emerging economies, smartphone and social media users have broader social networks. Available at: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/08/Pew-Research-Center_Emerging-Economies-Smartphone-Social-Media-Users-Have-Broader-Social-Networks-Report_2019-08-22.pdf. (Accessed June 1, 2023).
73. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: a systematic review of effects on patients and on their relationship with healthcare professionals. *BMC Health Serv Res.* (2016) 16:442. doi: 10.1186/s12913-016-1691-0
74. Suler J. The online disinhibition effect. *Cyberpsychol Behav.* (2004) 7:321–6. doi: 10.1089/1094931041291295
75. Ma X, Hancock J, Naaman M (2016). Anonymity, intimacy and self-disclosure in social media, Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems 3857–3869
76. Antheunis ML, Tates K, Nieboer TE. Patients’ and health professionals’ use of social media in health care: motives, barriers and expectations. *Patient Educ Couns.* (2013) 92:426–31. doi: 10.1016/j.pec.2013.06.020
77. Padrez KA, Ungar L, Schwartz HA, Smith RJ, Hill S, Antanavicius T, et al. Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department. *BMJ Qual Saf.* (2016) 25:414–23. doi: 10.1136/bmjqs-2015-004489
78. Fox S. (2011). The social life of health information. Available at: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf. (Accessed June 1, 2023).
79. Leyden S, Kolarova T, Bouvier C, Caplin M, Conroy S, Davies P, et al. Unmet needs in the international neuroendocrine tumor (NET) community: assessment of major gaps from the perspective of patients, patient advocates and NET health care professionals. *Int J Cancer.* (2020) 146:1316–23. doi: 10.1002/ijc.32678
80. AURORA Project (2018). The path to patient centricity: closing the “how” gap. Available at: <https://www.ipsos.com/sites/default/files/ct/publication/documents/2018-09/ipsos-healthcare-the-path-to-patient-centricity-august-2018.pdf>. (Accessed June 1, 2023)
81. ERG (Eastern Research Group) (2021). Assessment of the use of patient experience data in regulatory decision-making. Available at: <https://fda.report/media/150405/Assessment-of-the-Use-of-Patient-Experience-Data-in-Regulatory-Decision-Making.pdf>. (Accessed June 1, 2023)
82. Elstad E, Carpenter DM, Devellis RF, Blalock SJ. Patient decision making in the face of conflicting medication information. *Int J Qual Stud Health Well Being.* (2012) 7:18523. doi: 10.3402/qhw.v7i0.18523
83. Cuevas AG, O’Brien K, Saha S. Can patient-centered communication reduce the effects of medical mistrust on patients’ decision making? *Health Psychol.* (2019) 38:325–33. doi: 10.1037/hea0000721
84. Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. *Med Decis Mak.* (2015) 35:114–31. doi: 10.1177/0272989X14551638
85. Bonner C, Trevena LJ, Gaissmaier W, Han PKJ, Okan Y, Ozanne E, et al. Current best practice for presenting probabilities in patient decision aids: fundamental principles. *Med Decis Mak.* (2021) 41:821–33. doi: 10.1177/0272989X21996328
86. Matikainen M, Rajalahti T, Peltoniemi M, Parvinen P, Juppo A. Determinants of new product launch success in the pharmaceutical industry. *J Pharm Innov.* (2015) 10:175–89. doi: 10.1007/s12247-015-9216-7
87. Postmus D, Mavris M, Hillege HL, Salmonson T, Ryll B, Plate A, et al. Incorporating patient preferences into drug development and regulatory decision making: results from a quantitative pilot study with cancer patients, carers, and regulators. *Clin Pharmacol Ther.* (2016) 99:548–54. doi: 10.1002/cpt.332
88. Fayers PM, Machin D. *Quality of life: the assessment, analysis, and reporting of patient-reported outcomes.* 3rd ed. Hoboken, NJ: John Wiley & Sons Inc (2016).
89. Lohr KN, Zebrack BJ. Using patient-reported outcomes in clinical practice: challenges and opportunities. *Qual Life Res.* (2009) 18:99–107. doi: 10.1007/s11136-008-9413-7
90. Chang EM, Gillespie EF, Shaverdian N. Truthfulness in patient-reported outcomes: factors affecting patients’ responses and impact on data quality. *Patient Relat Outcome Meas.* (2019) 10:171–86. doi: 10.2147/PROM.S178344
91. EDPS (European Data Protection Supervisor) (2020). Preliminary opinion on data protection and scientific research. Available at: https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf. (Accessed June 1, 2023).
92. Kotsios A, Magnani M, Vega D, Rossi L, Shklovski I. An analysis of the consequences of the general data protection regulation on social network research. *Trans Soc Comput.* (2019) 2:1–22. doi: 10.1145/3365524
93. van Veen E-B. Observational health research in Europe: understanding the general data protection regulation and underlying debate. *Eur J Cancer.* (2018) 104:70–80. doi: 10.1016/j.ejca.2018.09.032
94. Moreno MA, Goniou N, Moreno PS, Diekmann D. Ethics of social media research: common concerns and practical considerations. *Cyberpsychol Behav Soc Netw.* (2013) 16:708–13. doi: 10.1089/cyber.2012.0334
95. Gonzalez-Hernandez G. (2019). On the ethics of using social media data for health research. Available at: <https://nlmdirector.nlm.nih.gov/2019/06/25/on-the-ethics-of-using-social-media-data-for-health-research/>. (Accessed June 1, 2023).
96. Tunçalp D, Lê PL. (Re)locating boundaries: a systematic review of online ethnography. *J Organ Ethnogr.* (2014) 3:59–79. doi: 10.1108/JOE-11-2012-0048
97. Fiesler C, Proferes N. “Participant” perceptions of Twitter research ethics. *Soc Media Soc.* (2018) 4:205630511876336. doi: 10.1177/2056305118763366
98. Golder S, Ahmed S, Norman G, Booth A. Attitudes toward the ethics of research using social media: a systematic review. *J Med Internet Res.* (2017) 19:e195. doi: 10.2196/jmir.7082
99. Bond CS, Ahmed OH, Hind M, Thomas B, Hewitt-Taylor J. The conceptual and practical ethical dilemmas of using health discussion board posts as research data. *J Med Internet Res.* (2013) 15:e112. doi: 10.2196/jmir.2435
100. Hemphill L, Schöpke-Gonzalez A, Panda A. Comparative sensitivity of social media data and their acceptable use in research. *Sci Data.* (2022) 9:643. doi: 10.1038/s41597-022-01773-w
101. Bazzano LA, Durant J, Brantley PR. A modern history of informed consent and the role of key information. *Ochsner J.* (2021) 21:81–5. doi: 10.31486/toj.19.0105
102. ESRC (Economic and Social Research Council) (2010). Framework for research ethics. Available at: <https://www.ukri.org/councils/esrc/guidance-for-applicants/research-ethics-guidance/framework-for-research-ethics>. (Accessed June 1, 2023)
103. Ford E, Shepherd S, Jones K, Hassan L. Toward an ethical framework for the text mining of social media for health research: a systematic review. *Front Digit Health.* (2021) 2:592237. doi: 10.3389/fdgh.2020.592237
104. Gelinas L, Largent EA, Cohen IG, Kornetsky S, Bierer BE, Fernandez Lynch H. A framework for ethical payment to research participants. *N Engl J Med.* (2018) 378:766–71. doi: 10.1056/NEJMs1710591
105. Beauchamp TL, Childress JF. *Principles of biomedical ethics.* New York: Oxford University Press (1979).
106. Pickering WG. Medical omniscience. *BMJ.* (1998) 317:1729–30. doi: 10.1136/bmj.317.7174.1729
107. EMA (2015). Good practice guide on recording, coding, reporting and assessment of medication errors. Available at: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guidelines-good-pharmacovigilance-practices-gvp-introductory-cover-note-last-updated-chapter-piv_en.pdf. (Accessed June 1, 2023).
108. EMA (2017). Guideline on good pharmacovigilance practices (GVP) module VI—collection, management and submission of reports of suspected adverse reactions to medicinal products (Rev 2). Available at: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-good-pharmacovigilance-practices-gvp-module-vi-collection-management-submission-reports_en.pdf. (Accessed June 1, 2023)
109. Convertino I, Ferraro S, Blandizzi C, Tuccori M. The usefulness of listening social media for pharmacovigilance purposes: a systematic review. *Expert Opin Drug Saf.* (2018) 17:1081–93. doi: 10.1080/14740338.2018.1531847
110. van Stekelenborg J, Ellenius J, Maskell S, Bergvall T, Caster O, Dasgupta N, et al. Recommendations for the use of social media in pharmacovigilance: lessons from IMI WEB-RADR. *Drug Saf.* (2019) 42:1393–407. doi: 10.1007/s40264-019-00858-7
111. FDA (2020). Patient-focused drug development: collecting comprehensive and representative input guidance for industry, food and drug administration staff, and other stakeholders. Available at: <https://www.fda.gov/media/139088/download>. (Accessed June 1, 2023)
112. FDA (2022). Patient-focused drug development: methods to identify what is important to patients guidance for industry, food and drug administration staff, and other stakeholders. Available at: <https://www.fda.gov/media/131230/download>. (Accessed June 1, 2023).
113. FDA (2022). Patient-focused drug development: selecting, developing, or modifying fit-for purpose clinical outcome assessments. Available at: <https://www.fda.gov/media/159500/download>. (Accessed June 1, 2023).
114. FDA (2023). Patient-focused drug development: incorporating clinical outcome assessments into endpoints for regulatory decision-making. Available at: <https://www.fda.gov/media/166830/download>. (Accessed June 1, 2023).
115. FDA (2022). Diversity plans to improve enrollment of participants from underrepresented racial and ethnic populations in clinical trials; draft guidance for industry. Available at: <https://www.fda.gov/media/157635/download>. (Accessed June 1, 2023).
116. Creswell JW, Plano Clark VL. *Designing and conducting mixed methods research.* 3rd ed. London, UK: SAGE Publications (2017).
117. Bodart S, Byrom B, Crescioni M, Eremenco S, Flood E. Perceived burden of completion of patient-reported outcome measures in clinical trials: results of a preliminary study. *Ther Innov Regul Sci.* (2019) 53:318–23. doi: 10.1177/2168479018788053