Social Media and M-Health Data

Subgroup report

Katherine Donegan, UK (subgroup lead)
Hans Ovelgonne, NL
Gavril Flores, MT
Per Fuglerud, NO (until October 2017)
Ada Georgescu, RO (until January 2018)

See websites for contact details
Heads of Medicines Agencies www.hma.eu
European Medicines Agency www.ema.europa.eu

An agency of the
European Union

Table of content

1. Summary

New and innovative technologies have led to an explosion in data coming from novel data sources that have not been widely utilized within medicines regulation. Social media is one such example that potentially brings a new dimension to healthcare as it enables more rapid and wider communication and data sharing directly to, from, and between patients than ever before. However, there are particular challenges raised by the form and availability of the data uploaded to social media platforms that need to be considered if the value of it to support medicines regulation is to be explored. It has been identified that it may have a complementary place within pharmacovigilance although further research is required to understand where it may compliment other sources of vigilance data. Further, the use of social media for direct communication with patients and healthcare providers is also being explored by medicines regulators and work is required to improve our understanding of how we can ensure effective communication that has a positive impact on public health.

mHealth technologies can provide access to large volumes of data that traditional data collection methods may not be able to gather potentially leading to more robust patient-centric evidence. The value of this within pharmacovigilance has already been demonstrated through a rapidly increasing number of individual studies and there is a need to continue efforts to understand how the advancing technologies can best enhance post-licensing surveillance. Validity of the data remains one of the key questions for medicines regulation when it comes to understanding how mHealth can support the pre-licensure phase while ensuring consistency of data collected in the post-authorisation setting is also important. There is also a clear need to understand how non-traditional endpoints coming from the use of such devices in clinical trials, such as those based on the use of accelerometers rather than established walking tests in patients with Duchenne muscular dystrophy, can be used to support clinically relevant and patient focused product information.

In both areas, it is recognized that regulators need to take a particularly collaborative approach given the rapidly evolving nature of relevant technologies and analytical methods and it is likely that there is a need to set out a clear road map, particularly for the use of mHealth technologies, for how regulators will tackle the challenges raised.

2. Background & definitions

2.1. Social media

The definition of "social media" is broad and constantly evolving. The term usually refers to Internet-based tools that facilitate the gathering of individuals and communities to communicate and share information, ideas, and experiences in real time.

Social media sites provide a variety of features that serve different purposes for the individual user. They can be grouped by purpose, serving functions such as social (e.g. Facebook, Twitter) and professional (e.g. LinkedIn) networking, media sharing (e.g. YouTube, Instagram), chat rooms and community forums (e.g. HealthUnlocked, WebMD Health Chat, patient.info, PatientsLikeMe), experience and opinion sharing (e.g. Tumblr), and aggregation or searching of information (e.g. Wikipedia, Google).

Social media has gained unprecedented worldwide popularity over the last two decades. It is estimated that there are currently over 2.3 billion active social media users internationally and this number is growing by approximately 1 million new users every day. General social networking platforms such as

Twitter and Facebook have built up a vast global reach while smaller networks and specialist forums facilitate the discussion of experiences around an issue or topic more rapidly, and with a wider more relevant network, than ever before. Unsurprisingly therefore, the use of social media to find, exchange, and discuss health information is growing at an unprecedented rate. Anyone with access to the internet can post or read information on a (social media) site. This means that it is directly accessible to patients, their family and friends, and all healthcare providers. Social media can provide healthcare professionals with tools to share information, to debate health care policy and practice issues, to promote health behaviours, to engage with the public, and to educate and interact with patients, caregivers, students, and colleagues. For patients, they facilitate fundamentally the same opportunities by providing access to active communities of healthcare professionals and fellow patients, with whom they can share information and their experiences, raise awareness of their concerns, learn about their conditions and healthcare opportunities, and find support.

By facilitating communication, the growth of social media and the internet has led to an explosion in data available directly from patients and healthcare providers. This potentially provides both a huge volume of information that could be of relevance for medicines regulation, which comes from a unique global perspective. However, these data are of quite a different type and format to that found in other established healthcare data sources such as clinical trials, registries, and electronic healthcare records, which raises a number of challenges. Further, social media platforms also provide a new and alternative route for regulators to communicate messages and data back to patients and healthcare professionals and facilitate a two-way discussion.

2.2. mHealth

Many of the advances that have driven the explosion in social media have also led to the discipline of mHealth (mobile health) which is the practice of medicine using technology such as mobile phones, tablet computers and other electronic devices, including wearables and implantable transmitting devices, to facilitate data collection to support self-management, clinical care, and research, and eventually to improve outcomes including through the use of digital interventions.

Smart phones are now ubiquitous with over 70% of adults in some EU countries reporting to have one. The capabilities of these devices have grown exponentially and telecommunications companies such as Apple have been very influential in the development of general mobile phone applications (apps) that enable the easy gathering and recording of a wide range of healthcare related data either directly inputted by the user or through passive collection using sensors (e.g. step-counters, heart rate monitors) as well as the analysis and presentation of the results for monitoring purposes.

General apps are widely used in the general population. For example, MyFitnessPal allow users to easily and exactly record their exercise and everything they eat and drink and provide them with the ability to monitor their nutritional intake over time, alerting them when a balanced diet is not being achieved. Further to more general apps, other specialist tools have been developed which allow the real-time recording of data directly by patients specifically related to their particular conditions as well as the rapid access to relevant information. For example, My Pain Journal allows patients with conditions such as rheumatoid arthritis and fibromyalgia to track their pain and symptoms through a specially designed pain diary adding details of the relief medication that was taken alongside any relevant photographs and then enabling reporting of them to a physician and recording them so that patterns can be explored. Another app, AsthmaMD reminds patients to use inhalers as scheduled, allows patients to monitor their peak flow alongside symptoms and triggers, and provides an action plan for asthma attacks. The anonymized data from this app is made available for researchers to help them correlate asthma with environmental factors, triggers, and climate change. Specific apps have also been designed for collecting data directly from patients within study settings.

There is some lack of clarity over what the term mHealth pertains to. While smart phones with apps and some basic sensors have certain uses, the scope of mHealth is much wider once the world of other electronic devices is considered. Specialist devices such as wearable motion detectors or blood pressure monitors, or more advanced implantable sensors that can gather and transmit data can also be included. Here, advances in technology may enable the gathering of vastly more data than ever before. The ability to exchange data through the growing connectivity of devices, the use of real-time analytics, and embedded systems has evolved the definition of the Internet of Things [ref]. This network takes us beyond the use of individual standalone devices. For the purposes of this report, the broader definition of mHealth will be used. As with social media the increased use of mHealth technologies enables the gathering of a large amount of data from new and unique sources. The value of this data for regulating medicines, facilitating access, and ensuring their safe and effective use should be considered.

3. Objectives

This report forms part of the work of the HMA/EMA Big Data taskforce. It is one of a set of mapping reports that cover a pre-identified range of Big Data sources capturing healthcare related data.

The specific objectives of this report are:

- To identify, characterise, and evaluate the data coming from existing social media sites and mHealth technologies that could be valuable to support medicines regulation decision-making.
- Use this mapping to identify stages in the product life cycle in which such data can facilitate or enhance regulatory decisions, and the challenges for regulators that raises, and to identify areas where it may potentially contribute but where limitations related to the data need to be addressed.
- To propose a set of recommendations to start to address some of these challenges facilitating the use of data, coming as it is from a rapidly evolving field, for regulatory purposes.

4. Methods

Given the very high number of potentially relevant social media sites and mHealth tools, a full mapping of all data sources was not feasible and so a general consideration of the characteristics of the data only is included. These characteristics are illustrated through a set of examples spanning the different types of sources considered potentially relevant. Social media and mHealth technologies are considered separately although there is overlap in some of the characteristics of the data and the issues raised. A literature search was conducted to support the identification of some of the potential uses of social media data and mHealth technologies within a healthcare setting where there may be regulatory implications. Additional relevant initiatives were identified through discussions within the regulatory network.

5. Data Characterisation

5.1. Social media data

5.1.1. Volume

As previously discussed, the number of people who engage with social media is vast and is still growing. Indeed, around 75% of the EU population has internet access and approximately 60% of those use social media, although only a proportion of these will be active users. This equates to approximately 400 million people across the EU. However, it is estimated that 90% of users are "lurkers", reading and observing social media but not actively contributing [van Mierlo, 2014]. The number of different social media sites also continues to grow. Therefore, the volume of data potentially available is vast although of course, only a small percentage of that will be healthcare related and only a very small proportion of that will be of relevance for any one issue.

Unlike more established healthcare data sources, the data arising via social media are predominantly unstructured free text, although there is some structured supporting data available. Social media sites are designed for the sharing of information not for the collection of research-quality data. To facilitate rapid communication, posts to social media sites are frequently concise in nature. They are also often isolated posts with limited details regarding the person who posted the message. This means that while there is variety in social media data, there is limited detail available. However, on some sites posts are linked together in conversations that increase the level of detail available. This is compounded by an increasing use of mobile devices to access social media sites, which encourage shorter posts with a greater use of abbreviations and slang terms. There is also extensive use of slang, abbreviations, colloquialisms, and images such as photos and emoji, which complicates analysis. These features have large implications when it comes to data analysis.

5.1.2. Veracity

Social media data come directly from each individual or group who engage with it. In the context of healthcare, this means data come from patients, their friends and families, and their careers, as well as healthcare professionals, regulators, and industry in a professional capacity. While an extremely large number of people use social media, as the choice to engage lies with the user it is important to consider how representative social media users are of the population, as this will have a large impact on the data available. There is a marked difference in the level of engagement with social media between younger and older people. Across the EU, around 75-90% of young people use social media compared to around 50% of the whole population. There are also differences across the level of interaction with social media and the sites actively used according to country, social class, and education level. Users will also engage with social media differently for personal and professional purposes. It should also be noted that social media has a global reach. While the direct relevance therefore to an EU population may vary, there also may be opportunities to use international data to inform decision-making. The population of users for any individual social media site may influence the data available from it. This may of course also change over time as the demographic distribution of the users changes. The 90-9-1 rule suggests that only 1% of users of social media contribute the vast majority of new content [van Mierlo, 2014]. Even then, of course, only information that a person chooses to upload to a social media site will potentially be available although it is though that 6% of internet users in the US have shared their own personal health experience online in the last 12 months [Fox et al. 2013]. Therefore, the completeness of the data for an individual person is unknown although it will be limited. Similarly, the representativeness of the data, or the patients contributing data, compared to the population and their experiences are also unknown.

As it is uploaded in real time by an individual or organization there is no validation or audit of the data. It must be remembered that user-generated content should be considered with a certain degree of caution as it may not be completely accurate or reliable and there may be considerable duplication. There is very little or no opportunity for researchers to influence the quality of the data. In addition, there is a need to be cautious of any information gathered from social media. Spam, malware, and phishing attempts are all circulated via such sites and may carry entirely erroneous data. Conversely, though it may be that people are more honest with the data they put onto social media compared to what they choose to tell a healthcare professional, which may mean the accuracy of the data in certain areas, is greater than that available in electronic healthcare records for example.

5.1.3. Variability

There is likely to be variation not only in the users of social media but also in the focus of the information, they post over time and across different sites. In the case of chat rooms and forums, this will be greatly influenced by the remit of the site itself. A large proportion of such forums are designed to bring together healthcare professionals from specific disciplines or patients affected by a certain condition. In the case of patient forums, in particular there is a trend towards a focus on serious and chronic or rare conditions and those predominantly affecting vulnerable populations. Posts on these sites will discuss issues potentially of interest to medicines regulation, including patient needs and priorities for treatment, their experiences, public perceptions of benefit risk issues, and potential adverse events, but will also discuss wider issues of health care that may hold less direct relevance to regulation. More general social networking platforms can potentially hold a much wider range of data than specialist forums as users can post whatever they wish. However, in turn that means that the vast majority of data will not even be healthcare related.

5.1.4. Velocity

Within Europe, there is already extensive use of social media. This means that while the number of people engaging with it is still growing, it is doing so at a slower rate than other regions in the world where current uptake is lower. However, the rate at which data are being accumulated is already rapid and is further accelerating. This has considerable implications for data handling and analytics.

5.1.5. Value

There are many features of social media data that will affect its value within medicines regulation. There has been extremely limited utilization of social media data in healthcare and therefore the potential uses are not clearly understood.

As described, issues of language, misspelling, colloquialisms and slang are all present within social media data. These features mean that, if the data are to be fully explored and potentially utilized, there is a clear need for algorithms and technologies to be developed in order to extract data from the source, identify the potentially relevant information, undertake data cleaning, and then conduct appropriate analyses. For example, for some research purposes, exposure and medical event/diagnosis dictionaries will need to be developed to identify relevant posts accounting for difference languages and errors in the text and link them to established coding systems algorithms for detecting duplicate records and anonymizing posts may be necessary. Machine learning approaches will be particularly of relevance. Of course, social media, particularly disease-specific forums, can be reviewed manually to increase general understanding or potentially used to identify or target specific patients but if the data are to be, robustly explored, then dedicated tools will be needed.

Other issues will also affect the value and use of social media data. These include the potential cost and difficulties associated with accessing the data, which will be held by private organizations, ethical, and data privacy issues related to the need for anonymization or potentially informed consent despite the public nature of the data, and the very limited opportunities for linkage to other healthcare related data sources.

5.2. mHealth data

5.2.1. Volume

As with social media, the number of people who engage with general mHealth mobile phone apps is vast and is still growing. There are currently more than 165,000 mobile health apps publicly available in major app stores, the vast majority of which are designed for patients [Cheng-Kai et al. 2017]. Specialist apps have also been designed for use in particular studies and trials. The size of the data coming from any one app will vary considerably. mHealth technologies are designed so that the data collected can be analysed and presented back. Therefore, data coming from individual mHealth devices are generally well structured although there may be issues with free-text data that the user is required to input, such as medication names for example, in terms of the readiness of the data for analysis.

5.2.2. Veracity

Data coming through mHealth technologies come direct from patients either through active engagement or through passive data collection. When considering mobile apps, the need for a smart phone or tablet computer means that, as with social media, the characteristics of users is not entirely representative of the whole population. However, providing the right technology directly to patients for their use in clinical trials or within routine clinical care can make use of more specialist health-related apps. Other devices will also likely have to be provided by healthcare providers or researchers although there is some use of simple wearable devices in the general population.

The quality and completeness of the data will vary according to the type of device and the setting in which it is being used. Devices that do not require input from the patient will likely have the greatest level of data completeness, assuming they are used all the time, which is not always the case, but are likely to be highly varied in terms of quality. Therefore, when they are used to gather data directly from patients within trials, data may not be complete however; it may be more complete than data from between study visits using more traditional methodologies. Further, while devices often collect process data (e.g. step counts) it is not always clear how this data is derived and different devices collecting apparently the same data may measure it differently and with differing degrees of this results accuracy. Furthermore, there is variability in access to granular raw data; some companies will provide access to raw data at its most granular level e.g. heart rate data organised in beat-beat intervals while other companies will not. This will affect the ability to validate the device and/or the evidence derived from it. General wearable devices such as Fitbits are now extremely popular within the general population with increasing numbers of more specialist wearable and implantable devices now available. Some of these can facilitate near real-time monitoring resulting in thousands of data points within an individual patient. However, there is variability.

5.2.3. Variability

Data deriving from mHealth technologies are likely to be extremely diverse and highly dependent on the type of device. When for example data originate from the log file of a pacemaker, the structure and quality can be expected to be clear and robust. In contrast, a mobile phone with a heart rate sensor is not a medical grade device and it will have been designed to be fit for a different purpose.

Likewise, readout from a continuous positive airway pressure therapy (CPAP) device will give precise measurements of apnea incidences. However, when data derive from an uncontrolled device that recreationally monitors sleeping quality and reports through a proprietary website into the cloud, there the quality and validity of the data are considerably less clear. Guidance has been produced by Medical Devices regulators within the EU on which mHealth technologies are considered medical devices, and hence require a CE mark in line with the EU Medical Devices Directive, and which is not [MHRA 2017]. In some instances, devices can be calibrated to individual users that may potentially reduce internal variation.

5.2.4. Velocity

The number of different mHealth technologies is growing rapidly. The speed of accumulation of data is likely to be lowest in a controlled setting in which a researcher sends a patient home with a device or where it is used within an RCT or observational study. Speed will probably increase as the measure of investigator-control is reduced in, for example, mobile applications that a patient can control and monitor independently and become highest in cloud-based applications where a whole group engages in an activity like losing weight.

5.2.5. Value

Additional verification of the data arising from individual mHealth technologies may be required depending on the purpose for which it is used. It is highly likely that data will be processed in an automated fashion, but to variable degrees and standards. Data from very well validated devices and apps are likely to be acceptable to support medicines regulation much more widely whereas data originating from less validated sources are likely to remain disputed and will have more restricted use. How to ensure that such data are sufficiently well validated for use in a regulatory context remains a question. The extent of validation needed to satisfy regulators will depend on the use of the data. For example, wearable devices used to support primary endpoints in pivotal clinical trials submitted as part of a licensing dossier requiring much greater validation in terms of the data they produce than apps used to gather data in pharmacovigilance studies.

There are likely to be opportunities for linkage of data gathered through devices and linkage with other data sources including electronic health care records. As with social media, outside of specific studies, data from mobile apps may be held by private companies and the availability of the data for research or to regulators will likely be restricted.

6. Examples of relevant data sources

In order to illustrate the complexities related to the data available via social media and mHealth technologies, and raised above, a set of examples have been chosen. The following examples have been chosen to reflect the spectrum of data sources. There were no systematic criteria used to select which data sources should be chosen for further discussion.

6.1. Examples of potential social media data sources

There are a considerably large number of social media sites that may hold potentially relevant data. One list of major social networking websites (Source: Wikipedia) which is not exhaustive and is limited to notable, well known sites includes over 200 different sites, around 30 of which have over 30 million registered users. However, the number of smaller patient forums and health-related communities is even larger. HealthUnlocked (https://healthunlocked.com/), for example, links over 500 communities, which are focused on conditions ranging from lifestyle factors such as healthy eating to rare conditions

such as Ehlers-Danlos syndrome. It also includes many forums dedicated to the discussion of specific drug substances. Separately, there are also a very large number of isolated forums that are run by specific charity groups.

The discussion above regarding the general characteristics of social media data can be illustrated and further explored by considering the two specific sites below in detail.

6.1.1. Twitter

Twitter (https://twitter.com/) is an interactive social media platform that was established in 2006. It allows account holders to post messages, `tweets', of up to 140-characters that are then shared either publicly or with other users who have chosen to `follow' them. Users are also able to upload photos and short videos and send private messages directly to other users. There are around 500 million tweets sent by more than 300 million active users worldwide every day. These can be made by an individual person or on behalf of an organization. Anyone can sign up to Twitter, and indeed a wide variety of people and groups have accounts, although younger people and those living in urban areas are more likely to engage with it.

The rate at which new users are joining Twitter is starting to slow although there are an additional 20 million active users now compared to early 2016. Approximately 10% of Twitter accounts are private meaning that tweets from that account can only be seen by followers of that account. Twitter also allows direct messaging from one account to another. These messages can only be seen by the two accounts involved.

Twitter users are not representative of the national offline population nor are they representative of all internet users. It is also important to note that Twitter data are not necessarily representative of Twitter users. This is because not all Twitter users will tweet on a topic of interest and so the data available is selective. It is also important to remember that it is not always individuals that may be tweeting but also, organizations, and those in a non-personal capacity, for instance journalists. Further to this it may even be difficult to ascertain whether a user or their tweets are real or fictitious. Anyone can set up a Twitter account and, whilst a mark of verification can be sought by the account holder, accounts and tweets are not routinely confirmed.

The size of this database means that there are likely to be data on the use and safety of medicines relevant to regulation [Sinnenberg et al, 2016 and 2017]. However, given that users can post anything they wish the vast majority of the data will not be at all related to health or medicines and so the challenge is to identify individual tweets that might be of interest.

As with other social media the information people make available on Twitter is free text. However, the restriction regarding the length of an individual tweet, and the facts that access to Twitter is often through a mobile device with tweets intended for sharing only with followers, means that the use of abbreviations and slang terms for example and misspellings are particularly common. Given the reach of Twitter the number of languages used is also vast complicating the use of the entirety of the database at once.

The nature of Twitter also means that posts are likely to be isolated and there is limited scope for identifying a breadth of information related to any one post. It is unlikely that there will be substantial healthcare related data for any one individual person and that a single mention of a healthcare-related issue will contain limited supporting data.

Twitter allows people to interact with its data, i.e. tweets & several attributes about tweets using Twitter APIs (Application programming interfaces). Twitter created open API allowing external developers to develop technology that rely on Twitter's data. The Search API can be used to query

tweets although limits are placed on these requests while the Streaming API pushes tweets out, again against a set of pre-defined criteria, as they happen. The only way to access 100% of tweets in real-time is through the Twitter Firehose API although it is associated with a significant cost. However, tweets from private accounts and direct messages are not included in this data.

While Twitter data can be used for research without informed consent, which would likely not be possible given the number of people and data points involved, there are ethical issues around reproducing tweets and consent to do so should be sought. Anonymization of the data should also be done prior to analysis to avoid individual identification.

The use of Twitter as a Tool for Health Research has been considered and there has been use of it specifically to explore the utilization of social media data for pharmacovigilance as discussed later in this report.

The majority of issues raised here with regards to the data available within Twitter are also relevant to other commonly used general social media platforms such as Facebook and Instagram. However, access to other general social media platforms varies. For example, in 2015, Facebook enacted a policy to no longer make post-level verbatim text available for data-mining purposes, citing privacy concerns. This effectively makes the world's largest social network unavailable for future research without special dispensation.

6.1.2. Inspire

Inspire: Health and Wellness Support Groups and Communities (https://www.inspire.com/) is an online patient community with over 1 million members who have contributed over 8 million posts. It brings together over 200 distinct forums for people affected by cancers, rare diseases, chronic conditions, and neurological disorders. These, often disease-specific, communities are developed in partnership with patient advocacy groups and user-generated content from these communities are in turn leveraged for secondary research by life science organizations. Around 65% of the members are patients while about 30% are caregivers. Inspire has over 700,000 monthly unique visitors.

There are number of potentially very relevant forums moderated by Inspire. For example, the Birth Control and Contraceptives community has several threads discussing adverse events potentially associated with different contraceptives while the Cancer Immunotherapy Community hosts discussions around patient experiences with patients sharing thoughts on their decisions and choices regarding treatment.

Inspire Insights™ offers clients' access to proportion of the data gathered within the patient forums enabling their use in secondary research. There are three degrees of privacy that Inspire users can select for posts published in an Inspire online community: (1) public posts that are visible to anyone (including outside visitors to the site), (2) public posts that are visible to any Inspire members, and (3) posts that are visible only to "friends" of author of the original post. Only 1% of Inspire posts are marked as "friends-only", thereby making the majority of posts publicly available. Acquisition of the data is dependent on compliance with Inspire's terms of use.

Many of the issues raised with Twitter previously are also relevant to data from Inspire. The form of the data is again free-text but unlike in more general social media sites the majority will now be related to healthcare issues and hence a larger proportion will be of potential interest for medicines research and regulation. Again, algorithms will need to be developed to identify relevant posts if the data are to be used quantitatively however, there are fewer restrictions on the length of posts and the organisation of posts into discussion topics means that individual posts will be richer and less isolated.

6.2. Examples of mHealth technologies

6.2.1. Wearable motion detectors

Physical activity and walking ability (distance, speed, quality) play a major role as potential patient-oriented outcome measures and confounding factors in a broad range of diseases; multiple sclerosis, coronary artery disease and chronic obstructive pulmonary disease being prominent examples. Insufficient assessment of this important variable may lead to unnecessary noise and even bias in the data, thereby diluting and/or diminishing a potential beneficial effect of a treatment. Very sensitive specialist motion sensors could also potentially be of value in tracking epileptic seizures, alerting of falls, or activity while asleep for example. Where previously studies have had to rely upon measurement of such activity only at study visits or through patient recall, the opportunities presented through wearable devices for more extensive or reliable collection of such data are increasingly being explored. Advanced wearable motion detectors are in use in various international multi-centre trials and clinical-epidemiological studies in, for example, multiple sclerosis, osteoporosis, Parkinson's disease, depression, and fracture heeling. Associated web platforms support central data management, analysis, and reporting and can often be directly linked to an eCRF or eTrial software. The devices measure, amongst other parameters, number of steps, gait speed, gait asymmetry, number of falls, and changes in altitude.

In comparison to some other healthcare data source, the data output from such sensors is reasonably well processed and complete as it is designed to be fed back to the user or healthcare professional. However, this is highly variable across different types of device and depends in part of the level of user interaction required. The main issues with using these data in a regulatory context are its unknown validity and hence its interpretation. Detecting subtleties in motion may be critical and even more reliable motion sensors may not be sensitive enough to provide sufficient data to allow the identification of the type of activity. Variability in the quality across different detectors is also an issue. However, the data will be readily available for research purposes and may be more robust that questionnaire type approaches for capturing long term data between study visits, particularly if motion sensors can be incorporated with GPS sensors to assist with interpretation of the data for example although such developments may cause issues related to patient confidentiality. Further improvements in the technology, which is evolving rapidly, may improve the validity of the data.

Unlike more complex motion detectors, simpler tools such as step counters would not be considered medical devices and as such would not have a CE mark or be regulated. The validity of the data is therefore likely to be poorer but to an unknown degree meaning it will likely have a complementary role compared to data collected using more traditional methods although may take a more prominent role in pharmacovigilance studies for example where again they may be more reliable than survey type methods.

6.2.2. Implantable diagnostics

Implantable devices can be used to monitor therapeutic measures such as blood glucose, coagulation, or creatine levels that can then provide information for calculating drug dose as well as drug level monitoring. Such devices have particular uses within chronic conditions such as diabetes, for example. Sensors placed below the skin can continuously monitor glucose levels with real-time results sent to a compatible mobile device, which can be used by a patient to determine when insulin is required. Such devices still require regular calibration using traditional finger prick tests.

Sensors such as these could play a critical part of personalised medicine. Advances in nanomaterial and in the technology needed to transmit the data are leading to increasingly complex biosensory devices used particularly in the treatment of chronic conditions although challenges remain in

harnessing robust data from them. They could also be paired with automatic or continuous drug delivery.

6.2.3. Symptom monitoring apps requiring user input

Currently, the top two categories of consumer-facing mHealth apps are wellness management (such as fitness, lifestyle modification, and diet and nutrition), and chronic disease management (such as mental health, diabetes, and cardiovascular diseases). The other categories include self-diagnosis, medication reminders, and electronic patient portal apps. Apps specific to physical medicine and rehabilitation are also available [Cheng-Kai et al, 2017].

Many clinical trials and observational research studies now make use of specialist apps for collecting data directly inputted by patients in between study visits. For example, the mPower: Mobile Parkinson Disease Study, which is using a specially designed app which implements both a number of surveys and tasks that activate phone sensors to collect and track health and symptoms of Parkinson's disease progression. The aims of the study are to learn about the variations of Parkinson's disease, to improve the way we describe and manage these variations, and to learn whether mobile devices and sensors can help measure Parkinson's disease and its progression to ultimately improve the quality of life for sufferers.

Patients who decide to join the study will need to download the study application on their own mobile device. Everyone who enrolls will first complete a consent process, explaining the risks and benefits of the study, and confirming their agreement to participate. An electronic registration will include entering their name, email address and other general information to verify your eligibility. Then periodically the patient will be asked to answer questions about health, exercise, diet, sleep and medicines, in addition to other surveys. One of the main issues for the use of this data within research is that a patient interacts with the app at their own convenience and may choose to participate in all or only in some parts of the study. The validity and completeness of the questionnaire data will therefore be very variable across different patients and will depend on their level of engagement with the app and the study.

Patients will also be asked to perform some activities via their mobile phone. The activities will be some brief tasks that need to be performed while holding the mobile phone like walking, taping or balancing for a short period. These passively collected data rely on the motion detector and sensors within the person's own mobile which will not have been designed for research purposes and hence the accuracy of the resulting data output might be poor.

In addition, if the patient is able to sustain moderate physical activity, they may receive motivational prompts to remain active. A unique random code will be associated with each patient's study data instead of their name. The coded data (without name and contact information) will be transmitted to the study team, added to the data of other study participants, and analyzed. If the patient consents the same coded study data can be made available to other qualified researchers for this and future research. In the future, the patient will have a unique account that they can use to review their own data. Such studies are going on in a wide range of clinical areas.

7. Potential regulatory applicability across the product life cycle

7.1. Social media data

7.1.1. Pharmacovigilance and identification of adverse event reports

Within medicines regulation, including within the EU as well as in the US FDA and Health Canada, there has been a focus on the use of social media within pharmacovigilance [Anderson et al, 2017, Curtis et al, 2017, Duh et al, 2016, Pierce et al, 2017, Powell et al, 2016, Price et al, 2016, Tricco et al, 2017]. Specifically, studies have explored the availability of comments made on social media that could help identify potential adverse events associated with a specific medicine. There is value in data reported directly from patients and the use of social media is likely to be much wider than that of passive spontaneous reporting systems. In order to explore this, new analytical techniques and platforms have been developed to identify potentially relevant posts [Cocos et al, 2017, Correja et al, 2016, Eshleman et al, 2016, Liu et al, 2016].

The Innovative Medicines Initiative and EFPIA co-funded a 3 year project which concluded in 2017, called WEB-RADR: Recognizing Adverse Drug Reactions, (https://web-radr.eu/) which set out in particular to provide access to social media data via a visualization platform to allow signal identification and confirmation as well as to develop and link new and existing analytical tools for the analysis of social media content for pharmacovigilance purposes. The consortium brought together expertise from world leading organizations across regulation, academia, the pharmaceutical industry, and technology companies and was structured to enable partners to participate in all areas, from shaping the regulatory framework that supported delivery of the project to participating in development and evaluation/research activities to ensure the products delivered offer the maximum possible benefits from regulatory, societal and scientific perspectives. In particular, the project built on work undertaken in part with the US FDA, which designed and implemented a social media listening and analytics platform, MedWatcher Social.

Research from this initiative, which is supported by smaller individual projects, showed that efficient semi-automated monitoring of social media sites, in particular Twitter, may provide earlier insights into certain adverse events although it was clear that such monitoring of social media will not replace current spontaneous reporting systems and manual review of potentially relevant social media posts is still required. This means that the potential value of the additional case reports needs to be carefully weighed against the burden required to access, analyze, and assess the data. It also suggested that some earlier research had been overly optimistic in its estimates of the performance of currently available methods for adverse event recognition in social media. The lack of specificity in general social media means that it is necessary to cast a wide net using a large collection of terms consisting of any symptom that could possibly be associated with the selected medical event when trying to identify potentially relevant social media posts. This results in a very large number of collected posts that ultimately are irrelevant but currently still require manual review although improvements in the natural language approaches used to identify relevant posts may reduce the extent of this. Similarly, it is likely to be difficult to identify the exact product used in many cases and this adds to the likelihood that there will be insufficient evidence within an individual post to mean it had any particular value. In addition, there is limited opportunity to follow up with the person who posted the data to verify or validate it or gather the further supporting information required due to the concise nature of posts to help assess potential causality. However, it is currently thought that social media monitoring may still be useful as an early warning system before an adverse event can be medically identified. There may also be particular scope in the data for identifying cases of misuse and abuse and medication error

[Anderson et al, 2017, Sarker et al, 2016]. Further outputs from the WEB-RADR initiative will be published shortly.

Further to the research being undertaken within WEB-RADR to explore the analytical methods surrounding the use of social media in identifying potential adverse events, additional aspects relevant to the use of social media within pharmacovigilance were also being considered. These included an assessment of personal data protection requirements in line with EU data protection legislation as well as ethical and societal aspects in relation to the use of social media for the purposes of pharmacovigilance and public health protection. The project has developed policy recommendations and input into regulatory guidelines and will describe the desired future, setting out what needs to be achieved in order to bring about change, in upcoming publications.

7.1.2. Social media listening

Several other uses of social media data have been identified and already considered within small studies in the literature. As already discussed, one of the key advantages of social media data is that it comes directly from patients and that they have actively decided to share this information. This can provide added value when compared to other data sources, which gather data predominantly from healthcare providers. This means that social media offer the potential to gather patient-centric data regarding their needs, perceptions, and priorities when it comes to healthcare and medicines that may differ from those envisioned by healthcare providers and regulators [Martinez et al, 2017, Topaz et al, 2016]. Daily problems and the burden of such experiences will be of high interest and importance to patients and therefore likely to be what they communicate on using social media while the interests of regulators and healthcare providers will be more focused on specific health risks and drug toxicity. Social media monitoring/listening is being conducted by industry and academia. An example of this is within vaccine pharmacovigilance, as there is considerable online debate and comments regarding the risks and benefits of vaccination per se and of individual vaccines [Becker et al, 2015]. While there are likely to be biases in the discussion, monitoring of social media for such issues may increase the understanding of public sentiment and the issues most of interest to them related to the use of medicines and could inform communication strategies. This data directly from patients could also help potentially provide signals related to drug effectiveness [Curtis et al, 2017, Risson et al, 2016] or issues related to prescribing practice [Hoang et al, 20160] and inform thinking regarding potential drug repurposing [0Rasetegar-Mojarad et al, 2016].

7.1.3. Use of social media for outgoing communication purposes

The reach of social media offers organizations, including medicines regulators who wish to get certain messages out to a diverse audience regarding the safe and effective use of medicines, a new route for direct communication. In addition, the widespread use of social media is also facilitating the spread of incorrect information among patients in particular and there is a need to engage to try to improve the quality of information accessed by patients.

In the UK, recent large-scale campaigns led by the MHRA on the use of sodium valproate in pregnancy and fake/counterfeit medicines have had considerable reach. Regarding sodium valproate, a paid for advert on Facebook reached more than 192,000 people, predominantly women aged 18-44 years while a further paid for Twitter ad reached over 12,000 people. Use of free-to-use social media including LinkedIn and Twitter have also reached a large number of people. A series of 14 tweets from the MHRA twitter account (https://twitter.com/MHRAgovuk) across a two-month period resulted in around 127,500 impressions, and over 400 and 650 likes and retweets respectively. Patient and charity organizations have been particularly supportive of our use of social media for this issue and have actively engaged with it themselves.

As part of the fake/counterfeit medicines campaign, a set of animations were produced and promoted on Twitter which resulted in 750,000 impressions and a 30% conversion rate of people clicking through to the MHRA website for further information. This campaign was further supported by television interviews and printed news stories and inclusion of a storyline in an ongoing drama series. Following the campaign searches of the UK register of legitimate online sellers increased by 33,706 (16% increase) and messages were estimated to have reached over 5million 18-30 year olds in total.

Social media has also been used within a coordinated campaign between national regulators across 21 EU member states as part of the Strengthening Collaboration for Operating Pharmacovigilance in Europe (SCOPE) Joint Action project for promoting national-level reporting from patients and healthcare professionals to spontaneous adverse event reporting schemes. The campaign led to a 13% increase in suspected ADR reporting (1,056 reports) between 15 regulators in the campaign week.

Successful social media campaigns likely benefit from simple clear repeated messaging, the use of graphics and pictures, and clear advice on action that the patient/healthcare professional should take. However, direct in-person interactions with different stakeholders has been important to increase the reach of messages through encouraging the resharing of regulatory messages on different organizational social media channels. The role of social media in sustaining such levels of engagement or supporting long term changes in clinical practice or patient behavior less unclear.

There is potentially further scope in exploring the use of social media and patient forums in particular for communication purposes as well as the use of social media as a way of engaging in a two-way conversation. While there are potential risks with this from a reputational perspective, particularly if two-way discussion is entered into, the presence of medicines regulators within such sites could help ensure consistency and accuracy of messaging and discussions.

7.1.4. Other potential uses of social media data

Monitoring social media also provides an opportunity to identify a potentially larger or wider cohort of patients who could be targeted to participate in further research, for example clinical trials or online surveys [Krischer et al, 2017, Moreno et al, 2017]. This could potentially support the identification of patients in whom recruitment to such studies would normally be low or facilitate the capture of data on a larger number of people. Of course, as with any sample regardless of how that is identified, consideration of the characteristics of the patients included and external validity remains important. Good Clinical Practice will also need to be adhered to if in a clinical trial setting. Whilst in a direct contact study setting there is scope to validate data and ensure a representative cohort of patients, in online surveys the characteristics of the responding cohort may be different to those who respond to surveys using other forms for example.

One use of social media data that has received particular attention outside of medicines regulation is disease surveillance and understanding. The use of social media to monitor infectious diseases and identify outbreaks has been considered by several researchers [Shin et al, 2016, Young et al, 2017]. However, studies have had mixed success. Researchers at Google explored the potential of services in their search engine to try and forecast the start and scale of annual seasonal influenza [Ginsberg et al, 2009, Sharpe et al, 2016], the principal idea being that when people start to become sick many will search on Google for related diagnostic and treatments options information providing almost instantaneous signals of flue prevalence by location. However, in the 2013 influenza season, the Google Flu Trends program failed to predict reliably predict prevalence and it was stopped [Lazer et al, 2014]. This demonstrated that such data should not be seen as a substitute for other data collection and that issues of validity and reliability remain, as the sources of these data are not designed to collect robust information amenable to scientific analysis. Smaller studies have also looked at trying to learn more about other conditions, for example depression and suicidal behaviors [Cavazos-Rehg et al,

2017, Mowery et al, 2017], and have suggested that approaches that are not designed to try to be so accurate may provide some useful more qualitative data.

Finally, social media of course provides the opportunity for direct patient to patient and patient to healthcare professional interaction to support self-management and management between appointments. This data could potentially complement the data already held on an individual patient within their healthcare record although linkage is of course extremely difficult and would likely need consent on a local level [Padrez et al, 2016].

7.2. mHealth data

7.2.1. Pre-licensing clinical trials & including patient reported outcomes

Within a regulatory context, data from disease management and symptom trackers apps and devices, including those making use of wearable or implantable sensors, may positively feature within pre-licencing clinical trials. In traditional trials, data are collected from study participants at specific pre-defined time points. However, continuous real time monitoring of symptoms and adverse events between study visits could be facilitated by introducing mHealth technologies. The use of specialist mHealth devices can increase the volume and completeness of between study visit data which is increasingly non-traditional and patient-centric and in the case of wearable and implantable devices, this can also increase the objectivity of such data.

The active engagement potentially involved in the use of mHealth technologies on the part of the patient where they are able to report outcomes directly may also help encourage increased study retention.

From a regulatory perspective, the validation of any data coming from such devices used in clinical trials submitted as part of a licensing dossier is highly important. Further understanding how endpoints coming from such trials which may differ from traditional endpoints can be validated, assessed, and used meaningfully in product information is also vital.

7.2.2. Pharmacovigilance and gathering spontaneous adverse event reports

In addition to the social media aspects, IMI WEB-RADR developed a mobile phone app designed to enable easy reporting of spontaneous adverse event reports directly from healthcare providers and patients to national competent authorities. This expands the prototype MedWatcher App platform already developed with support from the US FDA [Bahk et al, 2015]. In a specific study, and coupled with outreach via an online patient community, the MedWatcher app has allowed for rapid and more detailed individual case safety reports to be submitted with gains in efficiency. Three EU smartphone apps have now been launched through WEB-RADR by organization in the UK, the Netherlands, and Croatia. Analysis is underway on the use of the app versus conventional reporting including assessment of completeness and clinical quality of reports. These apps also provide regulators with a platform to alert users to new warnings including safety issues and allow the users to learn about adverse events that other users have experienced for products they have placed on their individual watch list. Within the project, the patient perspective is also being considered and the barriers and facilitators for use of the app are being explored. It has been found that two-way communication is a key factor in encouraging use of the app.

7.2.3. Implementation of risk minimisation measures and improving clinical care

Apps and other mHealth devices also potentially provide an opportunity for the implementation of risk minimization measures. This could include apps providing reminders to patients that could be hugely beneficial to patients in supporting them in controlling and optimizing their medication. Implantable devices ensuring that the correct dose is administered can also be linked to mobile platforms. However, again, here understanding how sensor technology data can be translated to meaningful and well-defined outcomes is very challenging.

There is also scope for integrating mobile health data into electronic medical records. Obviously, mHealth also enables the patient to take a much more active role in the management of their own conditions and enable more regular monitoring as well as providing an opportunity for healthcare interventions outside of a physical contact with a healthcare provider.

7.2.4. Post-authorisation safety and efficacy studies

As already discussed earlier in this report, apps have already been developed and are in use, for gathering data on symptoms directly from patients as part of real world studies. This highlights the existing value of such apps in post authorization safety studies. Similarly, mHealth measurement devices can also be used to generate real world data, which would be of value in post authorization studies.

8. Challenges with the use of these data in medicines regulation

8.1. Social media data

8.1.1. Analysis expertise within and outside the regulatory network

This review highlights a number of limitations of social media data that raise particular challenges for its use within a regulatory context. These data are very unlike the more traditional data types usually considered in healthcare research. The lack of structure poses one of the biggest challenges and this has major implications on the methodological approaches that need to be taken and the technology that is required in order to analyse it. Machine learning approaches, particularly natural language processing, are needed in order to extract and analyze relevant data from the vast amount of noise. Such techniques have, to date, only been seen within medicines regulation to a small degree and there is limited expertise within the network on their implementation. Regulators need to make sure that there are adequate links with researchers and organisations with the right scientific and technological expertise to explore innovative projects and new collaborations are explored. This expertise will largely be in academic organisations but collaboration with technology companies and data holders will also be required.

8.1.2. Requirements for further research into use in pharmacovigilance

It is clear that particular expertise is needed to develop and constantly test and improve the required methods and technologies as well as to interpret any findings and understand the limitations. IMI WEB-RADR proposed some clear recommendations for where further methodological research is required in order to further explore the use of social media for supporting the identification for adverse events.

This focuses on the identification of cases related to niche areas such as quality of life and abuse/misuse. Use of medicines in pregnancy may also be of interest.

A number of technical analytical methodology recommendations were also made. These relate to development of the machine learning approaches that need to be employed. This research needs to be supported by regulators to enable the development of internal expertise and to ensure that relevant guidelines are maintained or developed but as highlighted above the expertise will exist only outside the regulatory network.

There is also a need to understand any potential biases caused by using social media to identify patients for inclusion in post-authorization studies.

8.1.3. Data access & privacy

There are issues around access to the data, which although in many cases public, may need to be obtained from private organizations in order to be analyzed and the accessibility of this data to regulators and the pharmaceutical industry is not universal. The quality and reliability, or potential lack thereof, of this data also poses a challenge although other observational data sources are also potentially vulnerable to having invalid or biased information included. However, one of the biggest challenges is that these data are generally untested within healthcare research, so it is not understood if there is even potentially data of value within medicines regulation from social media sources beyond a suggestion of limited use within pharmacovigilance. Even in this particular field, research is still at relatively preliminary stages compared to that using more traditional data sources although the rate of research utilising social media data is rapidly accelerating.

Data privacy is a particular concern when using social media as while the information is freely shared by individuals there remains discussion about to what extent these data are available for research and if and when consent is required. These impacts upon the availability of data and, where it is available, upon the methods used for analysis.

8.1.4. Enhancing use in communication

As discussed, social media potentially brings a new dimension to healthcare as it enables more rapid and wider communication and information sharing between patients and their healthcare providers than ever before. It also provides medicines regulation and pharmacovigilance with a tool for regulators and the pharmaceutical industry to communicate with patients and healthcare professionals about the latest developments related to medicines and safety issues with the possibility of potentially improving health outcomes. However, there are challenges in understanding how to optimally use these newer routes of communication to ensure effective messaging resulting in improvements in public health, and the reputational risks of regulators engaging in active discussion on social media should be considered.

8.2. mHealth data

8.2.1. Understanding the use of mHealth in clinical trials

To promote the use of mHealth for research on medical treatments, the Duke-Margolis Center for Health Policy (DMCHP) collaborated with the US FDA to release a mHealth action plan that focused on the use of mHealth technologies for real world evidence generation [DMCHP FDA, 2017]. The plan offers substantial guidance on the use of mHealth in clinical trials, identifying opportunities for outcome

collection and the types of data that can be collected. A number of recommendations were made in that report which were designed to help create collaborations between different communities and mHealth companies, enable mHealth developers to build their products on a strong standardised base, and ensure that users of such technologies understand and can more efficiently consent to how their data is used. The role of regulators in this process needs to be fully established.

Mobile devices offer new ways to capture objective real time measurements from clinical trial participants including those related to both efficacy and safety although it needs to be ensured that these are being collected in a consistent way. This will enable the assessment of new endpoints that have not previously been possible to assess, or existing endpoints that can be measured in new and possibly better ways. These novel endpoints could provide high-quality data related to outcomes that are meaningful to patients while theoretically enabling larger trials with reduced barriers to participation making trials more sensitive and generalizable. However, how these endpoints can be identified, defined, and used in a meaningful way in patient and product information is not yet fully clear. The Clinical Trials Transformation Initiative (CTTI), a partnership between pharma companies, academics, and regulators, offered a set of recommendations for the use of mHealth in clinical trials specifically around developing novel endpoints using mobile technology in clinical trials [CTTI, 2017]. Understanding what the pathway is to wide acceptance of novel endpoints coming from such devices is important alongside agreement on what the role of regulators is in leading this.

8.2.2. Regulation of mHealth devices

There are also more technical barriers to using mHealth technologies. One of the challenges relates to those devices that are not subject to regulatory supervision [Cheng-Kai et al, 2017]. Governmental agencies, third-party companies, professional societies, and mHealth researchers have tried to come up with standards and systematic methods to evaluate and certify mHealth apps. There are also online app clearinghouses such as iMedicalApps (https://www.imedicalapps.com) that recommend apps based on editorial reviews.

There are also significant challenges raised even for those devices that are subject to regulatory supervision due to the pathway to approval and its governance structure in the EU. New mHealth devices are principally covered by the EU Medical Devices Directive (MDD) regulatory framework with additional legislation for some products within the Radio Equipment and Telecommunications Terminal Equipment (RTTE) Directive. In the EU, notified bodies, organisations designated by an EU country, assess the conformity of medical devices being placed on the EU market and the adherence of manufacturers to this directive. Manufacturers can apply to any single notified body in the EU and once they have the necessary certification, their products can be sold anywhere in, the EU meaning there is no centralised decision making regarding the approval of medical devices for use. The role of the notified body is to assess the manufacturer's quality system, examine the design and technical information relating to a product to ensure they meet requirements, either unit or batch test devices, and conduct audits of the manufacturers.

The role of national medical device regulators is then to ensure compliance of the notified bodies and manufacturers to any business activity relevant to the MDD and the General Product Safety Directive in accordance with the Consumer Rights Directive. They assess all allegations of non-compliance, monitor the activity of notified bodies, investigate medical devices as a result of adverse incident reports or other intelligence, and carry out projects to identify emerging risks, all as part of their market surveillance obligations.

Despite calls for moving to a single European regulatory body for devices, the series of EU-wide private organisations called notified bodies will remain following introduction of new medical devices legislation although there will be strengthening of the requirements on them to have the necessary expertise to

evaluate information supplied to them from manufacturers. To aid the notified bodies the European Commission will set up expert panels to advice on whether a high-risk device should be approved. This will be non-binding advice, but the panels' decisions will be published.

8.2.3. Regulatory requirements for validation and consistency of mHealth data

In the EU, as already mentioned, guidance has been produced by Medical Devices regulators on which mHealth technologies are considered medical devices and hence require a CE mark in line with the EU Medical Devices Directive and which are not [MHRA, 2017]. Even where a CE mark is required the level of regulation and the route of assessment, and therefore the data requirements, will differ according to the type of device. The classification of a device depends on the intended function, the duration of use, the invasiveness of the device, and the level of electrical activity. Class I devices can be self-certified while those with a higher risk classification require increasing levels of oversight by a notified body.

The most important challenge when considering the use of mHealth technologies to support the medicines product lifecycle, is that given the recent development of these technologies there is limited published literature citing the validity of the data arising from them even when they are regulated medical devices and as raised previously the level of validation required for devices to be placed on the market varies substantially. Therefore, it is important to understand exactly how mHealth technologies are regulated and to what degree and to be able to access any existing data on the validity of the device from device regulators or notified bodies where needed. Without this it is unclear if medicines regulators can be confident regarding the validity of mHealth devices and hence the robustness of the data arising from them for the proposed use.

European device regulation has been criticized in the past particularly with regards to the volume of data and rigor of testing required pre-approval and the extent and quality of risk management planning and post-marketing vigilance. The divided and somewhat fractured system of device regulation further affects the availability of data to medicines regulators to assess the validity of a device used alongside a medicine or in a research setting. Not all national regulators are responsible for the regulation of both medicines and medical devices. In some countries, these are separated and access to expertise on the regulation of mHealth technologies, and an understanding of the requirements for certification of specific mHealth devices, may be less easy to access so a coordinated approach to liaison with device regulators and notified bodies is required.

From a medicine regulatory perspective there is a clear need for evidence of validity if such devices are to be used to support medicines regulation and vigilance. If sufficient data demonstrating validity cannot be obtained from manufacturers, medical device regulators, or notified bodies, options for generating that data need to be explored. Given the rapidly evolving technology additional clinical trials examining the validity of such devices may not be appropriate, as by the time the trials are completed the technology may have moved on substantially. There are a range of different data types that will need validating e.g. the accuracy of measurements coming from motion detectors or the ability of patients to accurately report symptoms and use of medicines via apps used in pharmacoepidemiology studies. Given the spectrum of mHealth technologies there may be a need for not only large studies designed to broadly evaluate them, but also smaller but in-depth studies designed to evaluate specific approaches. The opportunities for development of standards for validation required for mHealth devices used to support medicines throughout their lifecycle, coordinating such studies, or establishment of an independent testing center/network, for example, could be explored after the specific regulatory requirements for validation had been mapped out.

Validating approaches to analysis of the data may also be needed to ensure they are separating noise from relevant data for some more complex wearable and implantable devices.

Concerns have been raised that too strict regulatory requirements could stifle the development and use of innovative technologies. For mHealth devices (i.e. those incorporating sensors etc.), the approach of regarding devices a black box which needs validation only from the outside may warrant some discussion. It is likely that some measure of understanding of the inner workings of the device and the nature of collected data will be of help in understanding how useful submitted data are for regulatory purposes. There may also be a need for regulators to consider how a balance between validating the reliability and robustness of the data coming from mHealth technologies whilst still encouraging engagement from the patient or user where such engagement is necessary.

Further technical challenges that still need to be addressed by the industry include issues around data privacy and security. Wearable devices could be lost or stolen and may be vulnerable to hacking. Therefore, mechanisms need to be in place for the safe storage and sharing of data enabling patient trust. Other technical challenges are the calibration of wearable devices to an individual and their characteristics and the potential misalignment of wearables, which affects their measurement quality and accuracy, over time. The durability and robustness of wearables may also require improvement if they are to be used in studies of a longer duration where sustained performance is required.

There is also inconsistency in the way that data are gathered by mHealth devices that are of relevance in their use in clinical trials but also post authorisation studies. For example, how drugs and conditions are captured and coded in apps where patients input medical history. Coding systems needed to support vigilance of devices are also less well developed than those for pharmacovigilance (e.g. MedDRA) which may present problems for monitoring and reporting on the performance of devices during longer-term studies.

8.2.4. Technical expertise requirements

mHealth is a relatively new and rapidly evolving area and it is clear that there is a need to work collaboratively to advance the use of such complex technologies in evidence generation. It is recognised that the relevant expertise to understand the full capability and workings of such devices does not sit within medicines regulators at present so identification of where the right expertise is, and initiation of a collaborative working environment are key to address questions on the validity of data, its place in clinical research, and how regulation must adapt. This expertise will likely be found in academic circles as well as in technology companies, including both small innovative companies and larger companies with portfolios of devices. This leads to challenges in understanding how potentially competing companies, and even possibly different academic groups should be engaged with, when they may benefit financially from increased use of their devices in practice or research and/or when medicines regulators are also regulating medical devices via notified bodies.

9. Recommendations

The mapping conducted in this paper and the subsequent discussion on the challenges and questions raised has been used to develop a set of proposed recommendations to support medicines regulators in the optimal use and understanding of data arising from these new and rapidly developing fields.

Given the lack of experience within regulation with regards to these areas there is likely need for a clear roadmap for moving forward and bringing together overlaps in the recommendations made below.

9.1. Social media

Topic	Core Recommendation	Reinforcing Actions	Strategic Goal
Pharmacovigilance and signal detection	Build on existing research on the use of social media data for providing insight into the identification of adverse event case report and to help understand how it can further be used to monitor the safety and effectiveness of medicines.	Focus on specific areas e.g. quality of life, exposure during pregnancy, abuse/misuse, to understand how social media may contribute useful data. Support further research into new analytical methodologies, including machine learning approaches to streamline the identification of relevant data. Investigation of a wider range of social media data sources particularly patient forums. Contribute to research on if and how social media reports can be integrated with other vigilance data sources. Actively promote a coordinated, transparent, and collaborative approach to future research in this field involving researchers and organisations with the right scientific and technological expertise.	To further develop an enhanced state of the art international pharmacovigilance system facilitating the rapid and robust identification of safety concerns.
Data access and use	Identify opportunities for gathering data from social media platforms.	Explore opportunities for regulators to access data from social media companies or to work with specific platforms to gather or stimulate new qualitative and quantitative patient reported data and take forward collaborations where appropriate. Ensure guidance is available and maintained on the ethical, moral, and legal implications of using social media data by the pharmaceutical industry taking the lead in its development.	Ensure regulators have access to relevant patient-centric data.
Patient recruitment into studies	Understand the implications of recruitment of patients to research through social media.	Start a project designed to establish the implications of recruiting patients to a range of different types of research study using social media on: • Ease of recruitment;	To improve the timeliness, robustness, completeness, and generalisability of data in clinical research relevant to

		• Robustness of data capture; • Follow up; • Generalisability. Do this with the aim of producing evidence-based guidance on the recruitment of patients into studies in this way.	the safe and effective use of medicines.
Communication	Actively research the use of social media for the communication of regulatory information.	Understand how behavioural science can contribute to effective messaging of regulatory recommendations on the use of medicines via social media to ensure changes in clinical practice. Measure impact of communications in a qualitative and/or qualitative way. Share experiences across the network on the use of social media by regulators for communication. Consider potential reputational risks and best practices for engaging in discussion on social media.	Support effective safety messaging, clinical management, and self-management.
Skills and knowledge within network	Equip regulators with the new skills required for this emerging area.	Ensure there is sufficient expertise within the regulatory network to identify and participate within relevant methodological research into the use of social media data within vigilance while recognising that academic and private organisations may be best placed to develop innovative approaches.	To optimally continue the exploration of the value of social media within medicines regulation and pharmacovigilance.

9.2. mHealth

Topic	**Core Recommendation**	**Reinforcing Actions**	**Strategic Goal**
Collaborative working	Bring together a group of relevant stakeholders including regulators, mHealth and pharmaceutical	Use this group to: Support learning within medicines regulators on technological capability, data	To promote the use of innovative mHealth technologies where they can support

	companies, academic researchers, and patient groups to advise on the use of mHealth technologies.	quality, analytical methodologies etc.; Help understand where mHealth technologies could have the greatest impact; Identify case studies that can be used to inform practice; Help develop best practice guidelines and establish where data are fit for purpose; Identify new challenges and areas for future focus; Feed into other mHealth workstreams on the use of the data in clinical trials and vigilance.	medicines development and their safe and effective use by actively engaging external expertise.
Medical devices regulation	Liaise with medical device regulators to ensure effective regulation of mHealth devices used to generate data submitted to medicines regulators.	Utilise existing routes of communication to engage with medical device regulators to ensure that the different regulatory frameworks can operate in a complementary way.	Ensure data quality and reliability are fit for purpose through effective and proportionate regulation.
Validation of data coming from mHealth devices	Facilitate the use of mHealth devices within research on the efficacy and safety of medicines.	Map the different types of mHealth data against their potential uses to define what extent and type of validation is required when from a regulatory perspective. Use this mapping to determine when specific guidelines are required and understand to what extent validation could be coordinated or how regulators could support independent testing.	Contribute to ensuring the quality of data submitted by more proactively defining expectations.
Use in clinical trials and patient reported outcomes	Ensure that medicines regulation is prepared for the submission of clinical trial data resulting from mHealth devices and facilitates its collection.	Review how mHealth data, particularly from wearable and implant type devices, have been used in clinical trials. Consider the need for specific regulatory guidance	Increase the clinical applicability and relevance to patients of the evidence available on the efficacy of medicines.

		on the use of novel endpoints arising from such devices in clinical trials. Agree on what extent of a role regulators should have in leading on the acceptance of novel endpoints.	
Collecting pharmacovigilance data and implementing risk minimisation	Support effective vigilance practices using state of the art mHealth technology.	Continue to develop apps for directly gathering data from patients on adverse events and encourage their wider use in real world and study settings. Do work to understand how apps and other mHealth devices might be used by patients to support risk minimisation and optimisation of their use of medicines and if yes where regulatory guidance on the collection, validation, and analysis of data is required.	Optimally explore the value of mHealth within pharmacovigilance.
Pharmacoepidemiology studies	Promote the use of mHealth technology to support effective post-authorisation studies.	Bring together case studies of where mHealth technologies are already being used in studies with active data collection to help understand where they could increase the strength of post-authorisation studies and the value of the resulting evidence to decision-makers and eventually patients. Work on ensuring standards for consistent data collection across apps are implemented.	Guide the use of mHealth within post-authorisation research.

10. References

Anderson LS, Bell HG, Gilbert M, et al. Using social media data to monitor misuse and non-medical use of bupropion: A content analysis. JMIR Public Health Surveill. 2017; 3: e6.

Bahk CY, Goshgarian M, Donahue K, et al. Increasing Patient Engagement in Pharmacovigilance Through Online Community Outreach and Mobile Reporting Applications: An Analysis of Adverse Event reporting for the Essure Device in the US. Pharmaceut Med. 2015; 29: 331-40.

Becker BF, Larson HJ, Bonhoeffer J, et al. Evaluation of a multinational, multilingual vaccine debate on Twitter. Vaccine. 2016; 34: 6166-71.

Cavazos-Rehg PA, Krauss MJ, Sowles SJ, et al. An Analysis of Depression, Self-Harm, and Suicidal Ideation Content on Tumblr. Crisis. 2017; 38: 44-52.

CTTI 2017. Developing Novel Endpoints Generated by Mobile Technology for Use in Clinical Trials. https://www.ctti-clinicaltrials.org/projects/novel-endpoints

Cheng-Kai K and Liebovitz DM. Consumer Mobile Health Apps: Current State, Barriers, and Future Directions. Clinical Informatics in Physiatry. 2017; 9: S106-15.

Cocos A, Fiks AG, Masino AJ. Deep learning for pharmacovigilance: recurrent neural network architectures for labelling adverse drug reactions in Twitter posts. J Am Inform Assoc. 2017 (epub ahead of print).

Correja RB, Li L, Rocha LM. Monitoring the potential drug interactions and reactions via network analysis of Instagram user timelines. Pac Symp Biocomput. 2016; 21: 492-503.

Curtis JR, Chen L, Higginbotham P, et al. Social media for arthritis-related comparative effectiveness and safety research and the impact of direct-consumer advertising. Arthritis Res Ther. 2017; 19: 48.

DMCHP FDA, 2017. Mobilizing mHealth Innovation for Real-World Evidence Generation. https://healthpolicy.duke.edu/sites/default/files/atoms/files/mobilizing_mhealth_innovation_for_real-world_evidence_generation.pdf

Fox S and Duggan M. Pew Research Center. Health Online 2013. http://www.pewinternet.org/files/old-media/Files/Reports/PIP_HealthOnline.pdf (accessed 25th July 2017).

Duh MS, Cremieux P, Audenrode MV, et al. Can social media data lead to earlier detection of drug-related adverse events? Pharacoepidemiol Drug Saf. 2016; 1425-33.

Eshleman R and singh R. Leveraging graph topology and semantic context for pharmacovigilance through Twitter-streams. BMC Bioinformatics. 2016; 6: 335.

Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. Nature. 2009; 457: 1012-4.

Hoang T, Liu J, Pratt N, et al. Detecting signals of detrimental prescribing cascades from social media. Artif Intell Med. 2016; 71: 43-56.

Krischer J, Cronholm PF, Burroughs C, et al. Experience with Direct-to-Patient Recruitment for Enrollment Into a Clinical Trial in a Rare Disease: A Web-Based Study. J Med Internet Res. 2017; 19: e50.

Lazer D, Kennedy R, King G, Vespignani A. The Parable of Google Flu: Traps in Big Data Analysis. Science. 2014; 343: 1203-1205.

Liu J, Zhao Sm Zhang X. An ensemble method for extracting adverse drug events from social media. Artif Intell Med. 2016; 70: 62-76.

Martinez B, Dailey F, Almario CV, et al. Patient understanding of the risks and benefts of biologic therpaies in inflammatory bowel disease: Insights from a large-scale analysis of social media platforms. Inflamm Bowel Dis. 2017. (epub ahead of print).

MHRA. 2017. Guidance: Medical device stand-alone software including apps (including IVDMDs) https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/610189/Software_flow_chart_Ed_1-03.pdf (accessed 18th June 2017)

Moreno MA, Waite A, Pumper M, et al. Recruiting Adolescent Research participants: In-Person Compared to Social Media Approaches. Cyberpsychol Behav Soc Netw. 2017; 20: 64-7.

Mowery D, Smith H, Cheney T, eta l. Understanding depressive symptoms and psychosocial stressors on Twitter: A Corpus-based study. J Med Internet Res. 2017; 19: e48.

Padrez KA, Ungar L, Schwartz HA, 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.

Pierce CE, Bouri K, Pamer C, et al. Evaluation of Facebook and Twitter monitoring to detect Safety Signals for Medical Products: An Analysis of Recent FDA Safety Alerts. Drug Saf. 2017; 40: 317-331.

Powell GE, Seifert HA, Reblin T. Social Media Listening for Routine Post-Marketing Safety Surveillance. Drug Saf. 2016; 39: 443-54.

Price J. What can Big Data Offer the Pharmacovigilance of Orphan Drugs? Clin Ter. 2016; 38: 2533-45.

Rasetegar-Mojarad M, Liu H, Nambisan P. Using Social Media Data to Identify Potential Candidates for Drug Repurposing: A Feasibility Study. JMIR Res Protoc. 2016; 5: e121.

Risson V, Saini D, Bonzani I, et al. Patterns of treatment switching in multiple sclerosis therapies in US Patients active on social media: Application of social media content analysis to health outcomes research. J Med Internet Res. 2016; 18: e62.

Sarker A, O'Conner K, Ginn R, et al. Social Media Mining for Toxicovigilance: Automatic Monitoring of Prescription Medication Abuse from Twitter. Drug Saf. 2016; 39: 231-40.

Sharpe JD, Hopkins RS, Cook RL, Striley CW. Evaluating Google, Twitter and Wikipedia as Tools for Influenza Surveillance Using Bayesian Change Point Analysis: A Comparative Analysis. JMIR Public Health Surveill. 2016; 2: e161.

Shin SY, Seo DW, An J, et al. High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. Sci Rep 2016; 6: 32920.

Sinnenberg L, Buttenheim A, Padrez K, et al. Twitter as a Tool for Health Research: A systematic Review. Am J Public Health. 2017; 107: e1-e8.

Sinnenberg L, DiSilvestro CL, Mancheno C, et al. Twitter as a Potential Data Source for Cardiovascular Disease Research. JAMA Cardiol. 2016; 1: 1032-6.

Topaz M, Lai K, Dhopeshwarker N, et al. Clinicians reports in electronic healthcare records versus patients concerns in social media: A pilot study of adverse drug reactions of aspirin and atorvastatin. Drug Saf. 2016; 39: 241-50.

Tricco AC, Zarin W, Lillie E, et al. Utility of social media and crowd sourced data for pharmacovigilance: a scoping review protocol. BMJ Open. 2017; 7: e013474.

Van Mierlo T. J Med Internet Res. The 1% rule in four digital health social networks, An observational study. 2014; 16: e33.

WHO. 2011. mHealth New horizons for health through mobile technologies. http://www.who.int/goe/publications/goe_mhealth_web.pdf (accessed 18th June 2017)

Young SD, Yu W, Wang W. Toward automating HIV Identification: Machine Learning for Rapid Identification of HIV-Related Social Media Data. L Acquir Immune Defic Syndr. 2017; 74: Suppl 2: S128-S131.