Commentary

What Can Big Data Offer the Pharmacovigilance of Orphan Drugs?

John Price, MB, ChB, PhD

Alexion Pharmaceuticals, Inc, New Haven, Connecticut

ABSTRACT

The pharmacovigilance of drugs for orphan diseases presents problems related to the small patient population. Obtaining high-quality information on individual reports of suspected adverse reactions is of particular importance for the pharmacovigilance of orphan drugs. The possibility of mining "big data" to detect suspected adverse reactions is being explored in pharmacovigilance generally but may have limited application to orphan drugs. Sources of big data such as social media may be infrequently used as communication channels by patients with rare disease or their caregivers or by health care providers; any adverse reactions identified are likely to reflect what is already known about the safety of the drug from the network of support that grows up around these patients. Opportunities related to potential future big data sources are discussed. (*Clin Ther.* 2016;38:2533–2545)

Key words: big data, drug safety, orphan drug, pharmacovigilance, rare disease, social media.

THE ORPHAN DRUG LANDSCAPE

Under the US Orphan Drug Act of 1983, a rare disease is one that affects <200,000 individuals (prevalence <650/million population);[1] ultra-rare diseases affect <20/million. In the European Union, Regulation EC 141/2000 defines orphan drugs as being for the treatment, prevention, or diagnosis of life-threatening or chronically debilitating disease affecting <5/10,000 (500/million).[2] Over a medical career, general practice physicians are unlikely to encounter an orphan disease, leading to a risk of missed cases even when an approved, effective therapy is available. It can take up to 20 years (average of 5.6 years in the United Kingdom and 7.6 years in the United States) to obtain a diagnosis for a rare disease, and patients with an orphan disease may visit a mean of 7.3 physicians before receiving an accurate diagnosis.[3] Many rare diseases are life-threatening or debilitating for the patient, emotionally and physically demanding for the caregiver, and financially burdensome to the health care system and society. Nevertheless, these so-called "orphan diseases" were historically an infrequent target for pharmaceutical companies because of limited opportunity for return on investment of expensive development costs, a lack of understanding of disease etiology, and the limited power of traditional drug discovery tools to target and modulate specific molecular targets. This situation has changed significantly in the last 10 to 20 years.

In the United States, the Orphan Drug Act of 1983[1] set up a dedicated office (ie, the Office of Orphan Products Development) and made provision for incentives for the development of drug treatments such as a period of market exclusivity and favorable tax treatment. The Rare Diseases Act of 2002[4] added the Office of Rare Diseases at the National Institutes of Health and increased funding, and draft guidance (*Rare*

Accepted for publication November 7, 2016.
http://dx.doi.org/10.1016/j.clinthera.2016.11.009
0149-2918/$ - see front matter

Diseases: Common Issues in Drug Development) was published by the US Food and Drug Administration (FDA) in 2015. Parallel initiatives, including the Precision Medicine Initiative[5] and the 2016 Orphan Products Natural History Grants Program,[6] encourage research into rare diseases and their treatment.

The EU Regulation EC 141/2000 established a centralized procedure for designating products as "orphan" and introduced incentives for the development of orphan products, including waived fees and market exclusivity for 10 years after authorization. The European Medicines Agency's Committee on Orphan Medicinal Products has reviewed applications for designation of a product as an orphan drug since its inception in 2000. Over the same time period, major scientific advances such as the Human Genome project and the eventual realization of the therapeutic promise of biotechnology, including monoclonal antibodies, antisense oligonucleotides, fusion proteins and many more approaches, have enabled pharmaceutical companies to bring forward potential drug treatments to address unmet medical needs for an increasing number of orphan diseases.

The result of these trends has been a large increase in orphan drugs under development and reaching marketing authorization in the United States and the European Union. There were 3735 products with orphan designation by May 2016, with 531 approvals since 1983[7] (Figure 1). There were 472 applications to the FDA for Orphan Drug Designation in 2015 alone, and the FDA's Director of the Office of Orphan Products Development, Gayatri R. Rao, MD, JD, predicted a 30% increase in 2016.[8] Of the 45 new molecular entities approved in 2015, a total of 21 (47%) were orphan drugs.

The European Medicines Agency website[9] listed 1314 products with active orphan drug designation by mid-2016 and an additional 373 products with a designation that had expired or been withdrawn; the number of orphan designation approvals by the Committee for Orphan Medicinal Products showed a steady increase from 2000 to 2015 (Figure 2). In the first 10 years after EC 141/2000, 58.3% of the marketing authorization applications for orphan drugs were approved,[10] and the number of marketing authorizations for orphan drugs increased 4-fold from ~3 per year to 12 per year over 15 years from 2001 to 2015.[11]

A comparison between the drugs approved in the first 7 years after the Orphan Drug Act and the drugs approved in the most recent full year (2015) listed at the FDA website reflects the evolution in science and increasing understanding of the molecular basis of

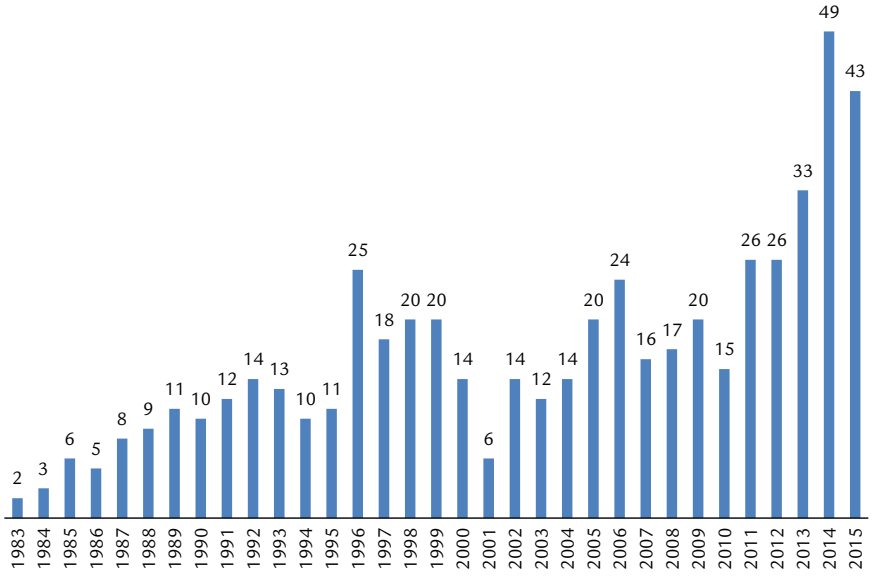

Figure 1. Approvals of drugs with an orphan designation by the US Food and Drug Administration, 1983 to 2015.

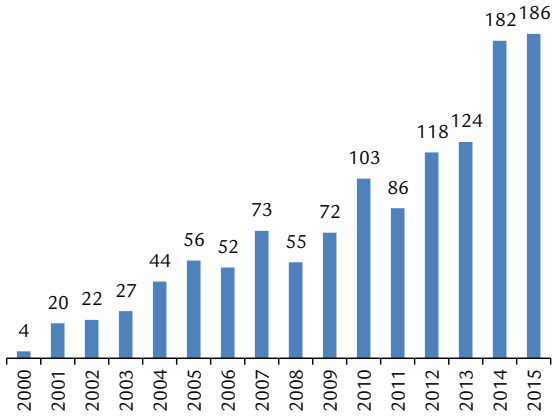

Figure 2. Number of active orphan medicinal product designations according to year of approval by the European Medicines Agency's Committee on Orphan Medicinal Products, 2000 to 2015.

rare diseases. Between 1983 and 1989, many orphan drugs had prior approval for non-orphan indications, some very common, and an "orphan" indication seemed opportunistic in targeting a thin slice of a larger population, such as late-stage fluctuations in Parkinson's disease, resistant malaria, and intravenous rifampicin for patients with tuberculosis unable to take the drug orally. Some of these early orphan drugs were a reformulation or new route of administration to enable patients to receive a product that was otherwise not available to them, even though their disease was not rare; for example, a lidocaine patch for chronically painful postherpetic neuralgia, intraspinal morphine infusion, and topical metronidazole for acne rosacea. Others (eg, thalidomide) were drugs that were no longer appropriate for wide use but showed value in treating a rare disease. For these products, knowledge of safety was previously well established.

The focus of the present review was on the challenge of pharmacovigilance of newer orphan drugs, intended to be used only in patients with a specific rare disease.

THE CHALLENGE OF PHARMACOVIGILANCE FOR ORPHAN DRUGS

Although the regulatory framework for pharmacovigilance does not differentiate orphan drugs, those orphan drugs that have been approved in recent years present unique challenges for pharmacovigilance departments. The limited size of a study population for an orphan indication, sometimes only a few dozen or 100 patients, results in a small adverse event dataset at the time of drug approval: for example, absence of an adverse reaction of interest among a total study population of 300 would only exclude, in accordance with the "rule of 3s," with 95% confidence, its occurrence as a "common" reaction.[12] Development clinical studies would often have been uncontrolled or comparative data might have been collected from a historical control group. Therefore, the safety profile, apart from common and predictable adverse reactions, is relatively poorly known.

Regulatory agencies have turned to seeking evidence to prove a hypothetical risk: for example, because a newly approved orphan drug is likely to be a biologic agent, administered parenterally, assessors have highlighted a risk of immunologic reactions. Experience suggests, however, that infusion-related reactions can exhibit unusual characteristics, including random and temporary occurrence after months or years of exposure that may not resemble the familiar allergic reactions to peanuts or insect stings. A new generation of orphan drugs targets lethal diseases of infancy and childhood, inborn errors of metabolism, and storage diseases; these very sick populations have a high "background noise" of adverse events related to the underlying disease. The transformative impact of a treatment that prevents death in infancy creates a new population of survivors who have a course of disease never previously seen, potentially surviving through adolescence to sexual maturity (including pregnancy while on therapy), to old age and the experience of diseases unrelated to the original diagnosis. A new adverse event detected as the underlying rare disease progresses while the patient grows older could be the consequence of the disease rather than an adverse drug reaction (ADR). The need to gather data postapproval is evident. Unsurprisingly, a review of 30 orphan drug approvals by the FDA between 2007 and 2009 found that 40% were approved with a postapproval commitment, and 17% required a postapproval study.[13]

Conventionally, for drugs administered to hundreds of thousands, or millions, of patients, a signal of a possible safety hazard arising from a single case report would be first evaluated by scrutinizing the worldwide safety database for similar case reports, as well as reports describing symptoms or signs or

laboratory findings potentially related to the events in the index case and review of evidence from clinical trial data in thousands of patients. The number of related reports of an event placed in context of drug exposure enables calculation of a reporting rate, which can be related to comparator drugs, or to occurrence of the event among untreated patients, to support assessment of causality. Applying mathematical methods such as Bayesian analysis[14] and calculating proportional reporting ratios[15] can automate, standardize, and speed detection of signals of drug-event pairs or show associations between concomitantly administered drugs and an event (potential drug–drug interactions).

However, for an orphan drug, there may be no other report of a similar or related event on a database containing safety data from a few dozen or few hundred patients. Common adverse reactions are those that occur in between ≥ 1 per 100 and < 1 per 10 exposures.[16] For a rare genetic disease such as a mucopolysaccharidosis, glycogen storage disease, or degenerative neuropathy affecting perhaps 1 per 100,000 children, the maximum population size for treatment in the United States and the European Union together might be much lower than 10,000 patients. With such a small population, application of the "rule of 3s" would mean that failure to detect an adverse event of interest among the treated population may only exclude, with 95% confidence, a risk greater than 1 per 3333 that the drug does actually cause that reaction.[12] Rare or very rare adverse reactions may not be detected. The low number of patients receiving an orphan drug has considerable significance to pharmacovigilance, and the new mathematical approaches to signal detection have limited applicability to small datasets.

Under these circumstances, a well-documented individual case report of an important suspected adverse reaction assumes great significance: strenuous effort should be made to gather comprehensive information about the patient and his or her treatment leading up to the event to establish causality, including temporal relationship, de-challenge and re-challenge, plausibility, potential risk factors, and alternative explanations. Accurate, comprehensive, high-quality information on individual cases is very important for the pharmacovigilance of orphan drugs.

Urgent pursuit of medical solutions to poorly understood, devastating rare diseases promotes self-help communities of impacted individuals with shared experiences; to support those afflicted with a rare disease, a close relationship forms between patients, families, caregivers, health care professionals, and drug manufacturers. Company-funded outreach patient support programs (PSPs) optimize use of the orphan drug by overcoming barriers due to the complexity of route of administration, the severity of the condition being treated, and drug access issues (eg, obtaining reimbursement). These programs are commonly run by nurses who educate the patient or caregiver about the disease and the orphan drug, promoting effective drug utilization and managing risk; the nurse may administer the drug. Adverse events collected through the PSP can provide a relatively large body of pharmacovigilance data, potentially mitigating the relative paucity of safety information that was available at the time of drug approval. This patient support can result in the overreporting of adverse reactions, with multiple reports of 1 reaction in a single patient received from different health care professionals involved in treating the patient, in addition to reports received from the patient.[17]

However, indiscriminate soliciting of adverse events, unfiltered by the responsible physician, from every patient (whether during or after the end of treatment) to satisfy regulatory expectations risks merely adding background noise to the safety database. Such data collection is paradoxically counterproductive to patient safety by inadvertently hindering signal detection and introducing inefficiencies that divert resources from more valuable pharmacovigilance activities. Opportunities exist for more thoughtful, targeted collection of high-quality safety data from PSPs, and disease or drug registries may offer preferable methods for systematic collection of safety data on orphan drugs. The ongoing revolution in biomedical informatics that fuels "big data" in life sciences by enabling the real-time collection of data on patients from wearable or implantable devices and monitors could augment this information in a meaningful way.

BIG DATA AS A SOURCE OF SUSPECTED ADRS

The digital revolution in communication has been accompanied by increasing interest from pharmaceutical companies, pharmacovigilance service providers, and drug regulatory agencies in evaluating whether information about drug safety might exist in big data. "Big data" is a popular term that may not have consistent meaning among all users. It may be described as a large volume of

data, structured and unstructured,[18] or a data integration strategy,[19] or in terms that infer dynamism and potential.[20] In addition to massive volume, characteristics of big data include variety (of data type and source) and velocity (of transmission and accumulation). Big data come from a large, often complex dataset on which traditional data-processing applications are inadequate; in biotechnological research, novel methods to enable intelligent review of new information focus on rapidly electronically identifying and associating potentially related discrete data points to overcome the logistic impossibility of detecting links between data points through a conventional manual search, now that the universe of data has effectively exploded in volume.[21]

Does big data offer a solution to the challenges of pharmacovigilance for drugs for rare diseases? The value of big data is predicated but also dependent upon novel, powerful analytical technologies yet to be realized.[20] Big data has the potential to determine new correlations between characteristics of patients within diverse datasets to create knowledge from information patterns, along with the expectation that future cognitive computing (and artificial intelligence) will reveal such associations more effectively than humans can.[22] Because the ability to identify and evaluate associations between drugs, events, and risk factors is the basis of pharmacovigilance, the appeal of big data is evident.

UTILITY OF SOCIAL MEDIA AS A TOOL IN PHARMACOVIGILANCE

Every digital transmission by e-mail, text, Tweet, Snapchat or Instagram photo, Facetime call, YouTube video, Skype teleconference, or website access generates big data, which accumulates continuously with unimaginable speed, volume, and diversity. Popular social media such as Facebook, Instagram, Snapchat, Twitter, and WhatsApp, individually or collectively, can be described as comprising big data, as well as chat rooms, blogs, and websites dedicated to specific diseases or patients. Previously undetected ADRs might be revealed by mining this unstructured data. Theoretically, searches of the content of these transmissions for combinations of words or characters commonly used to describe an adverse event could indicate discussion of an ADR, which could be confirmed by reviewing the online conversation returned by the search, provided that challenges including access and confidentiality can be overcome.

Certainly "Gen X", "Millennial," and "Generation Z" patients are increasingly likely to share their experiences of disease and treatments online, seemingly uninhibited by lack of data privacy, and it may be logical to think that important drug safety information could be found among online postings. Anecdotally, patients and caregivers have often turned to the Internet and online interactions to obtain drug safety and other medical information instead of (or to supplement) information obtained from their physician or pharmacist. Furthermore, proactively sharing an experience of drug use online could elicit instant feedback from other affected patients within a community of drug users, providing positive reinforcement for a patient's notions about a drug's effect.

Individuals intending to explore social media as a routine source of data for signal detection must acknowledge several fundamental drawbacks: (1) lack of specificity; (2) verification difficulties; (3) low validity; and (4) bias.

Lack of Specificity

A post of "feeling bad today. I took [X]" could indicate an ADR or that drug X was taken to alleviate symptoms. Also, to accommodate how different sectors of the population use the Internet, proprietary search methods are under consideration that include emojis (eg, "smiley faces" or "sad faces"), requiring an additional degree of creative interpretation of the search results.

Verification Difficulties

Follow-up confirmation that a patient genuinely experienced a reported suspected ADR may be difficult or unwelcome. It may even not be possible to identify the country of origin of the information being read, and establishing whether the suspected ADR satisfies the 4 minimum standard criteria for reportability may be hard.

Low Validity

Patients and caregivers without medical knowledge or access to diagnostic facilities are poorly placed to identify a medical condition, and they may be more likely to report symptoms and signs instead. Online conversations that include specific medical terminology could be rare, and information unfiltered by a health care professional merely adds background noise, hindering signal detection. A medical diagnosis posted online without medical confirmation is not

reliable evidence of an ADR. Future opportunities may emerge from telemedicine; this delivery of remote patient care through electronic communication between physicians and patients would create a new electronic source of medical information, including ADRs.

Bias

Data-mining social media carries an inherent bias toward those patients in socioeconomic, cultural, and age subpopulations who have technical ability and means of access. It may exclude elderly, vulnerable and infirm patients, who represent those using drugs and who are potentially at greater risk of ADRs.

Advantages and Disadvantages of Using Social Media as a Tool

Nevertheless, attempts are ongoing to identify suspected ADRs within patient discussions of drugs and events in the most readily available and accessible big data source (social media) by applying text-searching programs.[23–26] The FDA has signaled its interest to evaluate the potential of web search technologies to identify adverse reactions.[27,28] The EU WEB-RADR project reportedly collected and analyzed >1.5 million social media posts in English, Spanish, and French relating to ADRs by April 2016; the utility of the data collected was not discussed, however.[29] A related strategy analyzes logs of online searches made by Internet users.[30] Analysis of the search logs of 80 million Internet users detected disproportionate numbers of searches by users for terms of interest during a determined period of time after, compared with before, searches for a specific drug. The premise is that searches for drug information are proxies for drug exposure, and any search done after that time would have been made because the patient had experienced the adverse event.[31] The drawback of this assumption was evident when the program Google Flu Trends[32] wrongly signaled an outbreak of influenza when it detected a flood of flu-related searches by people who were concerned about press reports and were not themselves ill.[33] The searches done by the user are the equivalent of a medical information query rather than a report of an adverse event.

Despite the risk that self-reported experiences shared online are incomplete, inaccurate, or misinformed if the reporter is inexpert in medical concepts, the approach might be able to identify ADRs for widely used drugs earlier than reports appear on a regulatory agency safety database. However, the case reports identified seem relatively few, biased toward nonserious events, and appear to originate in younger patients, compared with reports gathered from conventional sources.[34] Terms identified from social media text-searches may be infrequently (<2%) confirmed to be adverse events, at a rate 5 to 10 times lower than potential adverse events detected by searching a structured company data source.[35] A few simple online searches confirm that for an orphan drug, social media is an even less promising source of adverse events than for widely used drugs. A search of Reddit and Twitter for pharmacovigilance information on 3 randomly chosen orphan enzyme replacement therapies used in children with inborn errors of metabolism, using proprietary software, returned 73 records mentioning the products over a 20-month period but no adverse events (personal communication, 2016), and a text search of Facebook for 1 of the 3 orphan drugs found 5 entries relating to its use, only 1 of which was related to a (labeled) adverse reaction (3 were complaints that the drug had not improved the patient's condition and 1 was related to a failed attempt to obtain the drug for a patient in North Africa). Even the preapproval clinical database for each of these orphan drugs would contain vastly more, and more useful, safety information than found online.

More promising online social media data sources may be websites dedicated to patient communities; patients often describe their personal medical experience, and many post their health records and laboratory data. A visual scan of the website "PatientsLikeMe,"[36] which includes an option for patients to evaluate their medications (including side effects), revealed that of the 1663 drugs with evaluations currently listed, the large majority are not for a rare disease, and there are many-fold fewer adverse event reports than are collected routinely by companies and regulatory agencies: ~270 products had been used by ≥100 patients (of whom approximately one third provided an evaluation), and 532 drugs had only 1 to 2 evaluations each. A further limitation of social media as a source of adverse events for drugs for rare diseases is difficulty obtaining follow-up information from the user because of data privacy issues, lack of contact information, and the patient's unwillingness to provide more information even when the user can be reached; some

reportedly find it "creepy" that an organization would be following their posts.[37]

Despite the significant limitations, there may be utility for social media as an early warning of an emerging safety issue: by soliciting similar drug use experiences from fellow drug users, a patient or caregiver might trigger a spike in reporting by other users with the same experience on a social media platform, within the community of drug users. Mathematical modeling of the relative contributions of different patient exposures, online communication behaviors, severity of risk, and frequency of risk needed to elicit such amplification for different ADRs could indicate the value of monitoring social media as an early alert system for all medicines.

PATIENT-SPECIFIC SOURCES OF BIG DATA

Big data might have been considered an appropriate description of the information available in databases of electronic health records (EHRs) (or electronic medical records). Analysis of a cohort of patients derived from these databases, who have the same disease or have been exposed to the same drug, can reveal an association between drug exposure and a health outcome, beneficial or adverse. Research conducted over the last 4 decades by epidemiologists working at or with organizations such as the UK General Practice Research Database, the Boston Collaborative Drug Surveillance Program, Saskatchewan Health, and similar databases shows that these datasets are amenable to existing analytical methods and have value in pharmacovigilance.[38–51] EHRs are created and maintained by medically trained personnel, a clear advantages over social media. Record linkage has not been widely adopted in pharmacovigilance, possibly because an observation of an association between 2 events could result from chance, bias, or confounding (rather than because 1 event caused the other) and therefore requires further investigation, including review of individual patient data[50,51]; this review can be laborious and costly to arrive at a definitive conclusion. In the United States, an additional factor is that the prerequisite for effective record linkage (ie, EHRs linked among medical facilities) has not existed consistently. The 2009 Health Information Technology for Economic and Clinical Health Act[52] provides incentives to encourage implementation of EHRs, and the Sentinel initiative[53] provides an opportunity for record linkage of US patients, focused on signal detection rather than confirmation.

However, record linkage methods may have limited application to pharmacovigilance of orphan drugs due to the paucity of patients with a rare disease in the population comprising the dataset and because of the isolated associations between drug exposure and events within that population. In investigating signals of drug–drug interactions identified from the FDA Adverse Event Reporting System database, EHRs provided additional information that strengthened a signal or enabled further evaluation of an association based on patient-specific clinical and laboratory data not recorded on this database.[54–57] However, the reported associations were not readily detected with less-used drugs from the same class, possibly because there were simply too few patients contributing experience with these drugs to the EHRs. Low patient numbers make EHRs a less attractive data source for suspected ADRs with orphan drugs. This limitation may be overcome, however, if EHRs can be adequately networked to create a big data source of multiple variables from multiple patients to make searching for new associations feasible.

A related data source is the company-sponsored PSP, from which it is normal to collect all adverse events regardless of causality or medical significance. Unfortunately, indiscriminate collection of all adverse events can result in the accumulation of events that are due to the underlying disease, nonserious events, and coincidental events of dubious relationship to the drug. For example, 85% of adverse events reported from Roche's PSPs were nonserious.[58] The value of the information gathered may also be limited, as seen by the review by Roche and the European Medicines Agency of 80,000 adverse events, including 15,000 deaths, in several PSPs: no new safety concerns were identified, the balance of benefits and risks of the medicines involved was not affected, and there was no new advice regarding product use.[59] Against a background of "noise," the potential to identify new safety information is limited by the conventional, human-directed technologies available but leaves open the possibility that hidden safety signals might be revealed with the future greater analytical power needed to evaluate big data.

FUTURE POTENTIAL OF BIG DATA: MOBILE DEVICES AND WEARABLE SENSORS

The versatile connectivity of modern mobile phones is an opportunity for pharmacovigilance. A US survey revealed significant interest among mobile phone owners in apps to track well-being, with approximately one half of smartphone users downloading health apps.[60] The user group was skewed toward a younger population and certain socioeconomic groups, and about one half of those surveyed had not continued using the app long-term; it seems that commonly available apps may tend to be used to track fitness rather than ill health. Apps can, however, be designed to collect data from patients with specific diseases or who represent specific patient populations and, depending on the data collected, could provide information that may correlate drug use with patient-recorded symptoms, signs, and clinical data.[61,62] Mobile apps are already available for use in clinical trials and, as long as the patient is "tech-savvy," offer several advantages over a pen-and-paper diary or scheduled on-line questionnaire; the investigator can track the patient's movement, trigger questions to the patient, readily collect patient-reported outcomes, and receive instant updates.[63–66] Collecting data through apps can deliver improved accuracy compared with pen-and-paper reporting by preventing data-recording errors and avoiding any opportunity for data entry errors on transfer to the database.[67]

Recognizing these potential implications, the FDA issued a guideline in 2015 on circumstances when such an app would be considered a medical device and subject to the applicable requirements.[68] In 2015, the Medicines and Healthcare Products Regulatory Agency implemented its own mobile app version of the "yellow card" reporting scheme for ADRs,[69] accessible by health care professionals and patients, in collaboration with the International Medicines Initiative Web-RADR; this action was followed in 2016 by Dutch[70] and Croatian[71] versions. Web-RADR is also currently surveying patients and health care providers to gather opinions about such technology.[72]

The use of mobile apps to record personal health-related experiences is therefore an emerging potential new source of pharmacovigilance data. A downside for pharmacovigilance of orphan drugs is, as mentioned earlier, that the patient's experience may already be under routine surveillance by the manufacturer, tracking a high proportion of treated patients, for example, through a registry or PSP. Patient-reported data collected through a personal mobile phone app is likely to differ qualitatively from the data obtained by the company from the health care provider involved in the registry or PSP, which may have already undergone medical evaluation (including an assessment of causality) and be reported along with other medically relevant data. If it were received by a company or agency today, an adverse event collected from a mobile phone app would be treated as a spontaneous report from a consumer and entered into the safety database to contribute to future signal detection and evaluation. Duplicate reports can be hard to eliminate if key information is missing, such as patient identifiers, or if different terminology is chosen by 2 reporters of the same event. If there is already a high level of reporting of adverse events, such mobile app data may not be additional but rather risk simply routing redundant, duplicate, and relatively low-quality information into the company's and regulatory agencies' existing pharmacovigilance systems, adding to background noise rather than amplifying a signal. The potential of mobile phone apps to contribute in a meaningful way to pharmacovigilance of orphan drugs may be in those instances where a registry or PSP is not available.

By contrast, data collected from wearable or implantable biosensors represent a novel, future source of information for pharmacovigilance for all drugs, presenting new opportunities. The concept of biosensors (devices that are capable of detecting the presence or concentration within the body of a biological substance through a receptor, antigen, antibody, enzymatic, or other detection element, and signaling the detection by electrochemical, optical, piezoelectric, acoustic, or thermal means) and the application of nanotechnology to biosensors[73–76] have opened the door to a future in which a patient's state of health and presence of disease markers can be monitored by wearable or miniature implantable devices transmitting data continuously to a portable device. Wearable or implantable biosensors are in use by patients with diabetes,[77] and an implantable glucose biosensor was recently approved in the European Union that works continuously for 90 days.[78] Today, we have the technology available to implant biosensors the size of microchips or smaller that monitor vital signs and changes in analytes and

detect the presence of biomarkers that might signify the presence of disease.[73,75,76,79,80] Although data received from these devices may currently go no farther than the personal receiver in the possession of the patient, a future state can be envisioned in which the data could be also connected and collectable centrally, thus contributing to big data. Linking patient databases to biorepositories (which could include biochemical or other parameters, or genomic information on individual patients) may create a data-rich environment within which to identify and explore correlations among populations of treated and untreated patients with a rare disease, including the identification of safety issues.[81]

Unfortunately, without enhanced computing capabilities promised by big data technology, the huge volume of data obtained from continuous patient monitoring, presented in its probable diverse formats, is likely to make it very difficult to correlate trends within the data with clinical observations about the patient and with the patient's drug use. If big data were to become merely another source of individual case reports of adverse experiences (eg, from mining social media), it would have little to offer pharmacovigilance of orphan drugs. It may even have a negative impact, if the discovered case reports were to lack the quality of reports of the same cases already received from health care providers and duplicate them unhelpfully while adding background noise that hinders signal detection. Companies and agencies depend on methods for detecting and evaluating signals that are largely human-directed. Conventionally, signals are identified by a manual or statistical algorithm that detects new or increasing prevalence of drug-event pairs, or drugA-drugB-event associations, in an accumulating safety dataset. Future ability to detect drug-event-event-event^n associations or event-drugA-drugB-drugC-drug^n interactions could enable earlier detection of previously unrecognized drug–drug interactions or syndromes associated with a drug (complexes of signs, symptoms and/or laboratory abnormalities [eg, serotonin syndrome,[82] "Drug Reaction with Eosinophilia and Systemic Symptoms"[83]]). Therefore, viewed not as a massive source of adverse events but as a means to correlate data within the set of data and identify previously unknown associations, big data may offer new possibilities for pharmacovigilance.

Although this view regarding big data is compelling, caution is needed. Massive datasets comprising multiple variables contain enormous numbers of potential intervariable correlations, risking the discovery of many falsely positive correlations by chance. The background "noise" of adverse event reports, which add little to existing knowledge of a drug's safety profile, might simply be replaced by a background noise of multiple correlations between variables. The risk of this outcome may be smaller for orphan drugs because low drug exposure and patient numbers would present fewer intervariable possibilities and a more manageable number of correlations to be validated when testing new technologies. Finally, however, once new technological capabilities have been established, a new challenge will present itself, that of tying new analytical capabilities into regulatory expectations and company accountability.

CONCLUSIONS

Pharmacovigilance of an orphan drug is dependent on receiving high-quality individual case safety reports of adverse events containing sufficient information to establish causality and risk factors because there are so few reports to analyze even if an ADR is "common." A background of severe and diverse symptoms due to underlying disease makes interpretation harder and calls for detailed patient information about each event. Even if the obstacles and biases of mining social media can be overcome, the low quality of data derived from this source restricts its value to signal detection for commonly used drugs with broad market penetration. For an orphan drug, it may even add background noise that hinders signal detection. Because of the close relationship between a patient with a rare disease and his or her caregivers, health care providers, and PSPs, duplicate reporting of suspected ADRs for orphan drugs occurs. Mining social media may merely reveal a duplicate (poorly documented) record of a suspected ADR that had already been reported by conventional pharmacovigilance. Sources of big data in which patients using orphan drugs are more highly represented, or which derive information directly from patients, offer greater potential.

Viewed merely as a source of adverse events, big data's potential for pharmacovigilance of orphan drugs may be overlooked. Given its size and complexity, its value may not be realized until technological advances have overcome the current limited scalability and bias of analyses done by humans. The real value

of big data is, therefore, in a situation in which a huge volume of information has accrued across large and diverse digital platforms, and analysis of that information with novel technologies has a reasonably greater probability of generating new knowledge than is already known. Methods to analyze big data are emerging, with a focus on creating logical associations between multiple discrete data points in datasets too large to analyze with conventional methods. Pharmacovigilance of orphan drugs may present a big dataset of manageable size in which to validate an emergent technology's ability to detect correlations between drug and clinical variables.

ACKNOWLEDGMENTS

The helpful comments of Paul Beninger during the writing of this review are acknowledged.

CONFLICTS OF INTEREST

None.

REFERENCES

1. US Food and Drug Administration, Orphan Drug Act—Relevant Excerpts: fda.gov/ForIndustry/DevelopingProductsforRareDiseasesConditions/HowtoapplyforOrphanProductDesignation/ucm364750.htm. Accessed 05.09.16.
2. European Medicines Agency, Orphan designation: ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000029.jsp&mid=WC0b01ac05800240ce. Accessed 05.9.16.
3. Engel PA, Sukir B, Broback M, Boice N. Physician and patient perceptions regarding physician training in rare diseases: the need for stronger educational initiatives for physicians. *J Rare Dis*. 2013;1:1–15.
4. The Rare Diseases Act 2002: www.congress.gov/107/plaws/publ280/PLAW-107publ280.pdf. Accessed 25.09.16.
5. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. 2015;372:793–795.
6. Federal Register, 2016: www.federalregister.gov/documents/2016/05/04/2016-10398/natural-history-studies-for-rare-disease-product-development-orphan-products-research-project-grant. Accessed 25.09.16.
7. FDA: accessdata.fda.gov/scripts/opdlisting/oopd/. Accessed 01.05.16.
8. Rao G, The Rise in Orphan Drug Designations: Meeting the Growing Demand, 2016, FDA Voice: blogs.fda.gov/fdavoice/index.php/2016/07/. Accessed 05.09.16.
9. EMA: ec.europa.eu/health/documents/community-register/html/orphreg.htm. Accessed 05.09.16.
10. Joppi R, Bertele V, Garattini S. Orphan drugs, orphan diseases. The first decade of orphan drug legislation in the EU. *Eur J Clin Pharmacol*. 2013;69:1009–1024.
11. European Commission. Inventory of Union and Member State incentives to support research into, and the development and availability of, orphan medicinal products — state of play 2015. COMMISSION STAFF WORKING DOCUMENT (2016), table 5, page 19. ec.europa.eu/health/files/orphanmp/doc/orphan_inv_cwd_20160126.pdf. Accessed 02.10.16.
12. Hanley JA, Lippman-Hand A. 'Rule of 3s' If nothing goes wrong is everything alright? *JAMA*. 1983;249:1743–1745.
13. Kesselheim AS. (2010) Innovation and the Orphan Drug Act, 1983-2009: Regulatory and Clinical Characteristics of Approved Orphan Drugs, in *Rare Diseases and Orphan Products: Accelerating Research and Development*, Appendix B.
14. Hauben M. A brief primer on automated signal detection. *Ann Pharmacotherapy*. 2003;37:1117–1123.
15. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidem Drug Saf*. 2001;10:483–486.
16. Guidelines for Preparing Core Clinical-Safety Information on Drugs. Report of CIOMS Working Groups III and V, Second Edition, Council for International Organizations of Medical Sciences (CIOMS), Geneva, Switzerland, 1999.
17. Drug Safety and Risk Management Advisory Committee Meeting, 2014: fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/DrugSafetyandRiskManagementAdvisoryCommittee/UCM426664.pdf. Accessed 05.09.16.
18. SAS. Big data: what it is and why it matters. sas.com/en_us/insights/big-data/what-is-big-data.html. Accessed 02.10.16.
19. Oracle. What is big data? oracle.com/big-data/index.html. Accessed 02.10.16.
20. IBM. Big Data. ibm.com/big-data/us/en/. Accessed 02.10.16.
21. Chen Y, Argentinis E, Weber G. IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research. *Clin Ther*. 2016;38:688–701.
22. Shaw J. (2014) Why "big data" is a big deal. Harvard Business review. harvardmagazine.com/2014/03/why-big-data-is-a-big-deal. Accessed 02.10.16.
23. Liu J, Zhao S, Zhang X. An ensemble method for extracting adverse drug events from social media. *Artif Intell Med*. 2016;70:62–76.
24. Liu X, Chen H. A research framework for pharmacovigilance in health social media: identification and evaluation of patient adverse drug event reports. *J Biomed Inform*. 2015;58:268–279.

25. Nikfarjam A, Sarker A, O'Connor K, et al. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *J Am Med Inform Assoc*. 2015;22:671–681.
26. Harpaz R, Callahan A, Tamang S, et al. Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug Safety*. 2015;37:777–790.
27. Comstock J. FDA taps PatientsLikeMe to test the waters of social media adverse event reporting. *Mobinews*. 2015. www.mobihealthnews.com/44366/fda-taps-patientslikeme-to-test-the-waters-of-social-media-adverse-event-reporting/. Accessed 30.10.16.
28. Kelley FA. *Google Searches Could Help FDA Identify Drug Side Effects*. 2015. www.newsinferno.com/google-searches-could-help-fda-identify-drug-side-effects/. Accessed 30.10.16.
29. Medicines and Healthcare products Regulatory Agency (2016) News story: Ground-breaking WEB-RADR project marks mid-point www.gov.uk/government/news/ground-breaking-web-radr-project-marks-mid-point. Accessed 30.10.16.
30. White RW, Tatonetti NP, Shah NH, et al. Web-scale pharmacovigilance: listening to signals from the crowd. *J Am Med Inform Assoc*. 2013;20:404–408.
31. White RW, Harpaz R, Shah NH, et al. Toward enhanced pharmacovigilance using patient-generated data on the Internet. *Clin Pharmacol Ther*. 2014;96:239–246.
32. Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457:1012–1014.
33. Butler D. When Google got flu wrong. US outbreak foxes a leading web-based method for tracking seasonal flu. *Nature*. 2013;494:155–156.
34. Duh MS, Cremieux P, Van Audenrode M, et al. Can social media data lead to earlier detection of drug-related adverse events? *Pharmacoepidemiology and Drug Safety*. 2016. online onlinelibrary.wiley.com/doi/10.1002/pds.4090/full. Accessed 30.10.16.
35. IMS Health (2015) Monitoring Adverse Events in Pharma's Patient Support Programs imsconsultinggroup.com/files/web/Global/Tech%20&%20Apps/Nexxus%20Commercial%20Application%20Suite/Nexxus%20Marketing%20Thought%20Leadership/Nexxus%20Social%20WP_Adverse%20Events_2015.pdf. Accessed 30.10.16.
36. PatientsLikeMe.com. Accessed 4.11.16.
37. Sukkar E. Searching social networks to detect adverse reactions. Pharmaceutical Journal. 294;7846 www.pharmaceutical-journal.com/news-and-analysis/features/searching-social-networks-to-detect-adverse-reactions. Accessed 04.11.16.
38. García Rodríguez LA, Pérez Gutthann S. Use of the UK General Practice Research Database for pharmacoepidemiology. *Br J Clin Pharmacol*. 1998;45:419–425.
39. Jick H. A major resource for drug safety studies. The General Practice Research Database. Carshalton: Centre for Medicines Research 1995.
40. García Rodríguez LA, Ruigómez A, Jick HA. Review of epidemiologic research on drug-induced acute liver injury using the General Practice Research Database in the U.K. *Pharmacotherapy*. 1997;17:721–728.
41. Castellsague J, Pérez Gutthann S, García Rodríguez LA. Recent epidemiological studies of the association between hormone replacement therapy and venous thromboembolism. *Drug Safety*. 1998;18:117–123.
42. Pérez Gutthann S, Garcí Rodríguez LA, Castellsague PJ, Duque OA. Hormone replacement therapy and risk of venous thromboembolism: population based case-control study. *Br Med J*. 1997;314:796–800.
43. Evans JM, MacDonald TM. Record-linkage for pharmacovigilance in Scotland. *Br J Clin Pharmacol*. 1999;47:105–110.
44. Van Herk-Sukel MP, Lemmens VE, van de Poll-Franse LV, et al. Record linkage for pharmacoepidemiological studies in cancer patients. *Pharmacoepidem Drug Saf*. 2012;21:94–103.
45. Malcolm E, Downey W, Strand L, McNutt M. West Saskatchewan Health's linkable data bases and pharmacoepidemiology. *PMS*. 1993;6:175–264.
46. Lewis JD, Schinnar R, Bilker WB, et al. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidem Drug Saf*. 2007;16:393–401.
47. Jick H, Vessey M. Case-control studies in the evaluation of drug-induced illness. *Am J Epidemiol*. 1978;107:1–7.
48. Jick H, Madsen S, Nudelman PM. Postmarketing follow-up at Group Health Cooperative of Puget Sound. *Pharmacotherapy*. 1984;4:99–100.
49. West SL, Savitz DA, Koch G, et al. Demographics, health behaviors, and past drug use as predictors of recall accuracy for previous prescription medication use. *J Clin Epidemiol*. 1997;50:975–980.
50. Furu K, Wettermark B, Andersen M, et al. The Nordic countries as a cohort for pharmacoepidemiological research. *Basic Clin Pharmacol Toxicol*. 2010;106:86–94.
51. Sørensen HT, Johnsen SP, Nørgård B. Methodological issues in using prescription and other databases in pharmacoepidemiology. *Norwegian J Epidemiol*. 2001;11:13–18.
52. HITECH Act, Enforcement Interim Final Rule 2009: www.hhs.gov/hipaa/for-professionals/special-topics/HITECH-act-enforcement-interim-final-rule/index.html. Accessed 05.09.16.

53. Ball R, Robb M, Anderson SA, Dal Pan G. The FDA's sentinel initiative—a comprehensive approach to medical product surveillance. *Clin Pharmacol Ther*. 2016;99:265–268.
54. Tatonetti NP, Denny JC, Murphy SN, et al. Detecting drug interactions from adverse-event reports: interaction between paroxetine and pravastatin increases blood glucose levels. *Clin Pharmacol Ther*. 2011;90:133–142.
55. Tatonetti NP, Fernald GH, Altman RB. A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. *J Am Med Inform Assoc*. 2012; 19:79–85.
56. Haerian K, Varn D, Vaidya S, et al. Detection of pharmacovigilance-related adverse events using electronic health records and automated methods. *Clin Pharmacol Ther*. 2012;92: 228–234.
57. Warrer P, Hansen EH, Juhl-Jensen L, Aagaard L. Using text-mining techniques in electronic patient records to identify ADRs from medicine use. *Br J Clin Pharmacol*. 2012;73:674–684.
58. Donzanti B. (2015) Evaluating adverse events from patient support and market research programs: proposed best practices and regulatory changes (oral presentation). 2nd Adverse Event Reporting and Safety Strategies Summit, Philadelphia.
59. European Medicines Agency, Press Release (2013) European Medicines Agency finalises review of medicines concerned by Roche pharmacovigilance inspection ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2013/11/news_detail_001962.jsp&mid=WC0b01ac058004d5c1. Accessed 25.09.16.
60. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. *JMIR Mhealth Uhealth*. 2015;3:e101.
61. Wahle F, Kowatsch T, Fleisch E, et al. Mobile sensing and support for people with depression: a pilot trial in the wild. *JMIR Mhealth Uhealth*. 2016;4:e111.
62. Woods SS, Evans NC, Frisbee KL. Integrating patient voices into health information for self-care and patient-clinician partnerships: Veterans Affairs design recommendations for patient-generated data applications. *J Am Med Inform Assoc*. 2016;23:491–495.
63. MetricWire (2014) Mobile Data Collection Apps: A Better Approach to Clinical Trials. metricwire.com/mobile-data-collection-apps-a-better-approach-to-clinical-trials/. Accessed 30.10.16.
64. Irissoftware Mobile App for Clinical Trials. irissoftware.com/upload/pdf/Iris_Success_Story_Mobile_App_for_Clinical_Trials.pdf. Accessed 30.10.16.
65. FierceBiotech Smartphone Apps for Clinical Trials. fiercebiotech.com/special-report/smartphone-apps-for-clinical-trials. Accessed 30.10.16.
66. Roche. Roche app measures Parkinson's disease fluctuations. roche.com/media/store/roche_stories/roche-stories-2015-08-10.htm. Accessed 30.10.16.
67. Zhang S, Wu Q, van Velthoven MH, et al. Smartphone versus pen-and-paper data collection of infant feeding practices in rural China. *J Med Internet Res*. 2012;14:e119.
68. FDA (2015) Mobile Medical Applications Guidance for Industry and Food and Drug Administration Staff fda.gov/downloads/MedicalDevices/…/UCM263366.pdf.
69. Medicines and Healthcare Products Regulatory Agency (2015) Digital evolution for ground-breaking Yellow Card Scheme. gov.uk/government/news/digital-evolution-for-ground-breaking-yellow-card-scheme. Accessed 25.09.16.
70. Web-RADR (2016) Lareb launch the Dutch version of the WEB-RADR app. web-radr.eu/2016/01/29/lareb-launch-the-dutch-version-of-the-web-radr-app/. Accessed 25.09.15.
71. Web-RADR (2016) HALMED WEB-RADR app launch. web-radr.eu/2016/05/19/halmed-web-radr-app-launch/. Accessed 25.09.16.
72. Web-RADR (2016) Mobile App Survey for Healthcare Professionals and Patients. web-radr.eu/2016/07/13/mobile-app-survey-for-healthcare-professionals-and-patients/. Accessed 25.09.16.
73. Wilson GS, Gifford R. Biosensors for real-time in vivo measurements. *Biosens Bioelectron*. 2005;20:2388–2403.
74. Vaddiraju S, Tomazos I, Burgess DJ, et al. Emerging synergy between nanotechnology and implantable biosensors: a review. *Biosens Bioelectron*. 2010;25:1553–1565.
75. Vaddiraju S, Legassey A, Wang Y, et al. Design and fabrication of a high-performance electrochemical glucose sensor. *J Diabetes Sci Technol*. 2011;5:1044–1051.
76. Bohunicky B, Mousa SA. Biosensors: the new wave in cancer diagnosis. *Nanotechnol Sci Appl*. 2010;4:1–10.
77. Koschwanez HE, Reichert WM. In vitro, in vivo and post explantation testing of glucose-detecting biosensors: current methods and recommendations. *Biomaterials*. 2007; 28:3687–3703.
78. New Implantable CGM Receives Approval (2016) diabetesincontrol.com/new-implantable-cgm-receives-approval/. Accessed 30.10.16.
79. Fracchiolla NS, Artuso S, Cortelezzi A. Biosensors in clinical practice: focus on oncohematology. *Sensors*. 2013;13:6423–6447.
80. Ferguson BS, Hoggarth DA, Maliniak D, et al. Real-time, aptamer-based tracking of circulating therapeutic agents in living animals. *Sci Transl Med*. 2013;5:213ra165.
81. Rubinstein YR, Groft SC, Bartek R, et al. Creating a global rare disease patient registry linked to a rare diseases biorepository database: Rare Disease-HUB (RD-HUB). *Contemp Clin Trials*. 2010;31:394–404.

82. Sternbach H. "The serotonin syndrome.". *Am J Psychiatry*. 1991;148: 705–713.
83. Cacoub P, Musette P, Descamps V, et al. The DRESS syndrome: a literature review. *Am J Med*. 2011;124: 588–597.

Address correspondence to: John Price, Global Pharmacovigilance and Drug Safety, Alexion Pharmaceuticals, Inc, 100 College Street, New Haven, CT 06510. E-mail: john.price@alexion.com