Addressing the Data Challenges of Pharmacovigilance

INTRODUCTION

Increasing data volumes as well as increasing data complexity are currently forcing the drug safety industry to look for solutions to reduce case processing costs while remaining compliant with continually changing regulations worldwide, as well as maintaining or even improving the quality of information contained in individual case safety reports. As pharmacovigilance adopts next-generation technology by leveraging artificial intelligence (AI) and the cloud, new possibilities are opening up for knowledge generation – and thus value – from the data collected and processed.

Cognitive computing has been changing the world significantly. Many industries have been employing such technologies effectively and efficiently for some time now. The adoption of cognitive computing has also started in the medical sciences yet key questions about how to integrate it successfully into the various global healthcare systems need to be addressed. In order to achieve successful integration, there is a need to overcome technical and medical limitations, as well as regulatory obstacles. In addition, ethical concerns, in particular regarding the safety and integrity of data, need to be resolved. Several areas of medicine are unthinkable without the application of cognitive computing and artificial intelligence, for example personalized medicine. The sheer depth and wealth of data required for streamlining patients to allow them decision-making on the optimization of their personal health benefit-risk balance in light of their health state, their lifestyle choices, genetic and metabolomic factors, and preventative and curative medications, require advanced computational approaches. At the same time, the tendency to oversell the technology needs to be curbed and a rather realistic, step-wise approach should be followed to successfully implement and optimize cognitive computing and artificial intelligence approaches in order to really meet the needs of stakeholders.

One of the cornerstones for the optimization of the benefit-risk balances of drugs has been the proactive management of drug safety risks through the systematic collection of information about adverse events associated with drugs in real-life clinical settings through pharmacovigilance. The pharmacovigilance world, however, has been traditionally slow in adopting concepts from other industries or sciences – for example, formal risk analysis methodologies like FMEA that have been used in other industries for decades were introduced to pharmacovigilance processes only less than ten years ago, and proactive risk management thinking has replaced reactive PV in Europe only after the 2012 PV legislation came into force. The amount of data collected from markets as well as from dedicated post-authorization studies of the safety of medicines, however, in combination with an increasing number of additional data sources that have become available over the last years, necessitates the switch to AI applications also for pharmacovigilance.

Besides the pressure coming from the amount of data the market generates, there is also regulatory demand for the integration and management of ever larger amounts of data that are being used in the processing and evaluation of drug safety information. The EMA's drive to move towards the ISO IDMP (Identification of Medicinal Products) standards to establish definitions and concepts and describe data elements and their structural relationships requires significant adaptations for industry stakeholders to meet these standards. The process to implement IDMP in order to integrate information from and for pharmacovigilance, regulatory submissions, clinical trials and Good Manufacturing Practices has been keeping industry busy – and will continue to do so. The adoption of E2B(R3) standards for regulatory reporting is only one aspect of these ongoing changes.

Artificial intelligence in adverse event processing

Concepts like the application of artificial intelligence (AI), through techniques like natural language processing (NLP) and deep or machine learning (ML), have started to affect the pharmacovigilance world. Adverse event processing is one of the most obvious targets for the automation of PV processes as this has been a repetitive task, routinely performed by all pharmaceutical companies and usually a rather significant cost item in the PV continuum. Just over half of respondents, 57% according to a recent survey of PV experts, have in the last years outsourced the labor-intensive task of manual adverse event case processing to dedicated companies (*See exhibit 1*). The processes, however, have not changed – adverse event case information enters the workflow, needs to be coded according to the requirements of the respective company's case management system and adverse event database, as well as the terms laid out in the dictionary for the coding of adverse drug reactions (MedDRA). There is a step for the adjudication of relatedness, i.e., whether the adverse event reported could plausibly be caused by the suspected drug (or a combination of that drug with other reported co-medications) and the medical assessment of the event's seriousness. Finally the case enters the database in the (hopefully) correct format, ready to be reported as an electronic file in E2B format to the regulatory authorities according to the regulatory reporting requirements in all the countries where the drug is marketed. In most companies with a significant number of marketing authorizations active in more than just a few countries, the latter task is also automated, based on a table of reporting requirements that takes into account the characteristics of the case.

It is, we believe, quite obvious that a large part of the case management process therefore lends itself to automation and some process steps (like electronic case submission) have already been automated. That automation has certainly led to an increase in the quality of submissions compared to the manual submission of forms that had been the standard before E2B standards became the norm. Besides, as electronic submissions occur quasi-instantaneously and at the same time to all recipients, there are no delays in regulatory reporting timelines to be expected from these steps.

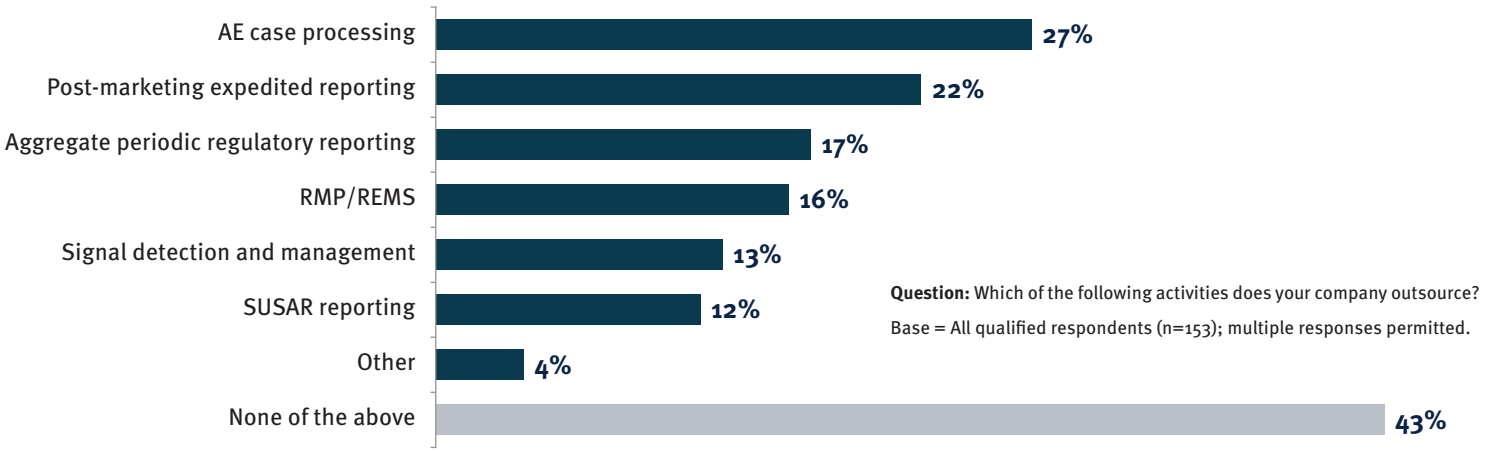
It can thus easily be envisioned that the most significant impact to be expected from further AI applications in PV is on the quality and the speed of the work, compared to manual performance of tasks that can be automated. Routine, repetitive tasks such as adverse event processing can definitely be largely automated in a sophisticated way as also exemplified by a recent workshop of the International Society of Pharmacovigilance (ISoP, December 2017, Boston, MA) that indicated that the time is ripe for the adoption of the respective machine tools. We can envisage that case information obtained in natural language from reporters can be dissected automatically by appropriate automata to deduce the relevant information on patient, suspected drug, adverse event, and reporter from the text. Given the development of learning algorithms we can also foresee that the creation of case narratives should be possible as well as the simple determination of the listedness or expectedness of an adverse event. For newly discovered adverse events there will still be a need to employ human medical expertise in order to determine the potential causality of an event, mak-

OVER **60%** of respondents are implementing or planning to implement AI-supported technologies for case processing in their companies.

Exhibit 1

Activites Outsourced

Just over half of respondents (57%) outsource at least some activities, most commonly AE case processing and post-marketing expedited reporting.



Question: Which of the following activities does your company outsource?

Base = All qualified respondents (n=153); multiple responses permitted.

ing the human part of the overall process the scientifically and medically most significant and interesting.

Automation in case processing may, however, mean different things to different organizations. Not all companies will require or desire the same extent of automation. This will depend on various aspects, like the complexity of case management processes within the respective companies, case volumes, or the case processing workflow steps, to name a few. Most of the currently ongoing projects in case management automation are likely focusing on simplifying case intake and triage as these initial workflow steps lend themselves to a carefully built automation framework, making some of the manual steps obsolete. A combination of natural language processing and machine learning concepts can be used for the simplification of the case intake and triage processes. Going a step further, it may be conceivable to use optical character recognition (OCR) technology, allowing for self-reading of incoming source data and differentiating information contained therein. The hurdles to be overcome to achieve this, however, should not be underestimated. Everybody who has worked with the complex and various structured or unstructured source data, or had to differentiate lay and medical terminologies to identify a patient's medical history and the actual reported adverse event from these data knows about the significant challenges in the process. If software engineers and pharmacovigilance experts, however, closely work together to develop suitable algorithms and train the automata on a variety of source data with increasing complexity, these challenges can be dealt with.

Given the technical feasibility of case processing automation, we should be able to witness the emergence of increasingly refined and complex automation tools that will be further improved by exposing them to the real-world challenges of case management, resulting in well-tested and validated systems, reducing the time required for case processing while improving the consistency of case quality, and ultimately resulting in significant efficiency gains. While we certainly will not see a complete elimination of human input into the case management process, in particular in the application of medical and scientific knowledge, we can expect a massive reduction of manual labour in such routine tasks that will free up pharmacovigilance experts to focus on strategic tasks, such as signal detection or benefit-risk assessment and management. Naturally, the increase in efficiency will also lead to a decrease of per-case-cost in the process.

It will be interesting to see how comfortable biopharmaceutical organizations will be in adopting such tools. The survey (*See exhibit 2*) revealed that over 60% of respondents are implementing or planning to implement AI-supported technologies for case processing in their companies. Most prominently, quality checks, follow-up processing, expedited reporting and medical assessment are ranking on top of the tasks to be automated. Most of these have been mentioned above and can be considered to lend themselves to automation. Medical assessment, on the other hand, is a rather surprising response and may need to be further investigated as the complexity of this process step is rather high and it is questionable whether initial medical assessment of relatedness or the evaluation of seriousness criteria from the narrative is meant rather than the de novo assessment of drug exposure – adverse event causality.

Now what would be the biggest hurdles to overcome when deciding whether or not to go for automation of PV? Most respondents (*See exhibit 3*) cited a lack

Exhibit 2
Implementing Artificial Intelligence In AE Case Processing

The majority of respondents (62%) are actively implementing, or planning to implement, artificial intelligence in AE case processing. The areas in which they are most likely to do so include quality check, follow-up processing, expedited reporting and medical assessment.

Question: In which areas of AE case processing are you actively implementing or planning to implement artificial intelligence?

Base = All qualified respondents (n=153); multiple answers permitted.

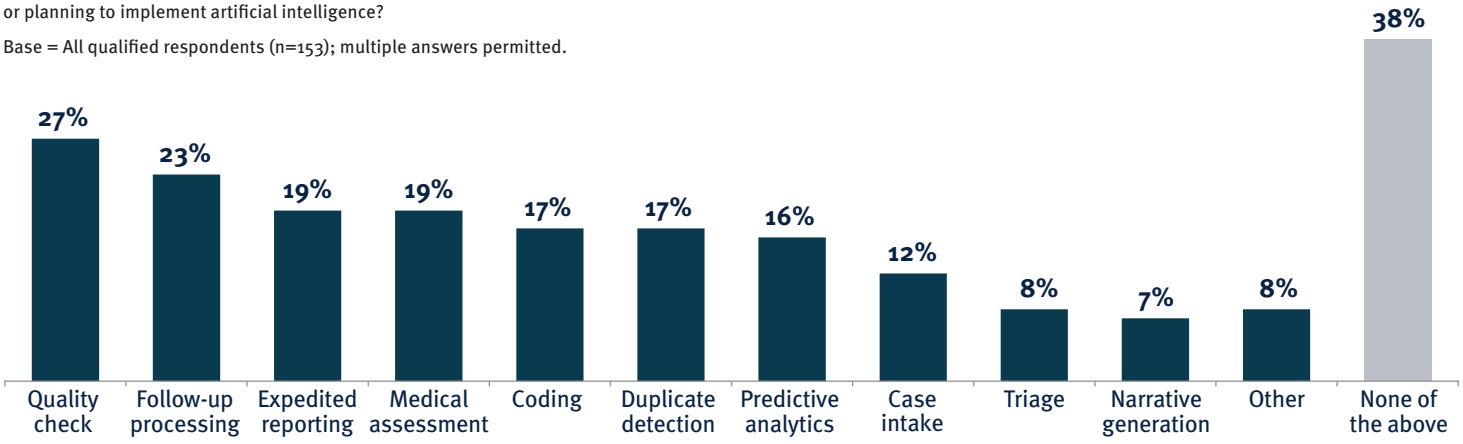
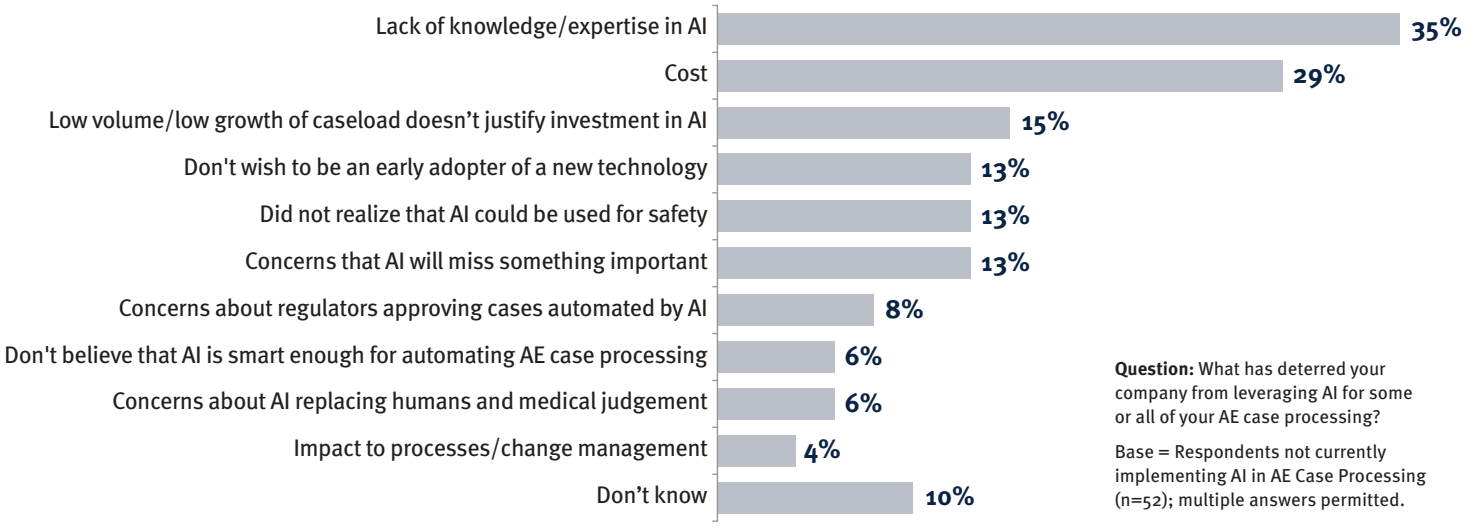


Exhibit 3
Deterrents To Implementing AI In AE Case Processing

Respondents not currently implementing AI in AE case processing report lack of knowledge in AI and cost are key deterrents.



Question: What has deterred your company from leveraging AI for some or all of your AE case processing?

Base = Respondents not currently implementing AI in AE Case Processing (n=52); multiple answers permitted.

of knowledge in AI and expected costs. Related to the lack of AI knowledge, it was also interesting to see that a number of respondents did not realize that AI could be used for safety. The lack of knowledge can certainly be tackled by appropriate training and education efforts, and learned societies (e.g., ISoP) have been making efforts to explore the knowledge space in this area and will hopefully continue to do so, as well as providers of software solutions for these tasks. As mentioned above, there will be a significant need for software engineers and PV SMEs to cooperate very closely on these tasks and to learn from each other in order to come up with solutions that are not only technically working but also do not expose companies to potential compliance-related challenges that could occur if tools are not sufficiently tested or validated. Thus, all technological advances, as bright as they may seem, need to carefully balance the technological possibilities against the regulatory requirements for case processing and the applicable pharmacovigilance regulations.

Costs associated with automation are a rather surprising response as the adoption of AI tools can – as has been shown in other industries – be safely assumed to reduce cost in a very short term. The cost pressure in many companies may be the reason why investments into new technologies are delayed, although they could bring a relatively quick return on investment and an immediate gain in efficiency.

Overall, it can be said that the time is definitely ripe for educating the undecided about the benefits of PV automation and to develop this area, keeping in mind that automation concepts should be developed through collaboration between leading safety database providers, specialized pharmacovigilance service providers and biopharmaceutical organizations, including, wherever possible, direct input from regulators to ensure the adoption of tools and methods that are meticulously planned, thoroughly validated, and implemented while maintaining stringent compliance requirements.

Cloud solutions for pharmacovigilance

Many industries have been benefiting from the ability to store and analyze huge amounts of data in the cloud. As more and more data sources are contributing to the integrated knowledge of benefits and risks of medicinal products in various clinical settings, the requirement to optimize the intake, storage and analysis of these data becomes more urgent for the biopharmaceutical industry as well. It is obvious that pharmacovigilance could benefit from the ability to build a massive and robust database of diseases, medicines, and adverse events that would ultimately be able to allow all parties, from physicians to patients to research institutions, globally to access all data around authorized drugs and their side effects.

Medicine today, and by inference the pharmaceutical industry, are limited by an incomplete understanding of the biology of disease. Imagine a massive database incorporating all that is known about metabolic pathways, signaling pathways, the interactions between enzymes governing these pathways, genetic and epigenetic factors influencing them, diseases related to the factors involved in the pathways, medicines affecting them – independent of whether this occurs directly through on-target or off-target effects – and associated health outcomes. Add to this the integrated knowledge about pre-clinical and clinical study data, pharmacoepidemiological data on real-world outcomes, patient-reported outcomes, patient preferences

about health states, as well as regulatory opinion, and one arrives at a universe of knowledge about diseases and the medicines intended to treat them. Such a vision may have been completely utopian only a few years ago as the storage requirements for the data associated, including the wealth of new data added every instance, would have been impossible to implement. With the emergence and further development of methods for data intake into the cloud and the management and analysis of these data within the cloud, the vision comes closer to reality. We already see how more and more data such as those mentioned above are stored and handled in various cloud architectures and how systems' pharmacology-based approaches to perform mechanism-based pharmacovigilance are being investigated. System networks are being generated by integrating data (and more importantly knowledge) from various sources. Academic researchers from Stanford University recently presented a network-based a priori algorithm for association mining in FDA FAERS reports, thus expanding pharmacovigilance from a reactive mode to the mechanism-based prediction of adverse drug reactions.

58%

of respondents either have some or all of their safety solutions in the cloud or are planning to move there within the next two years.

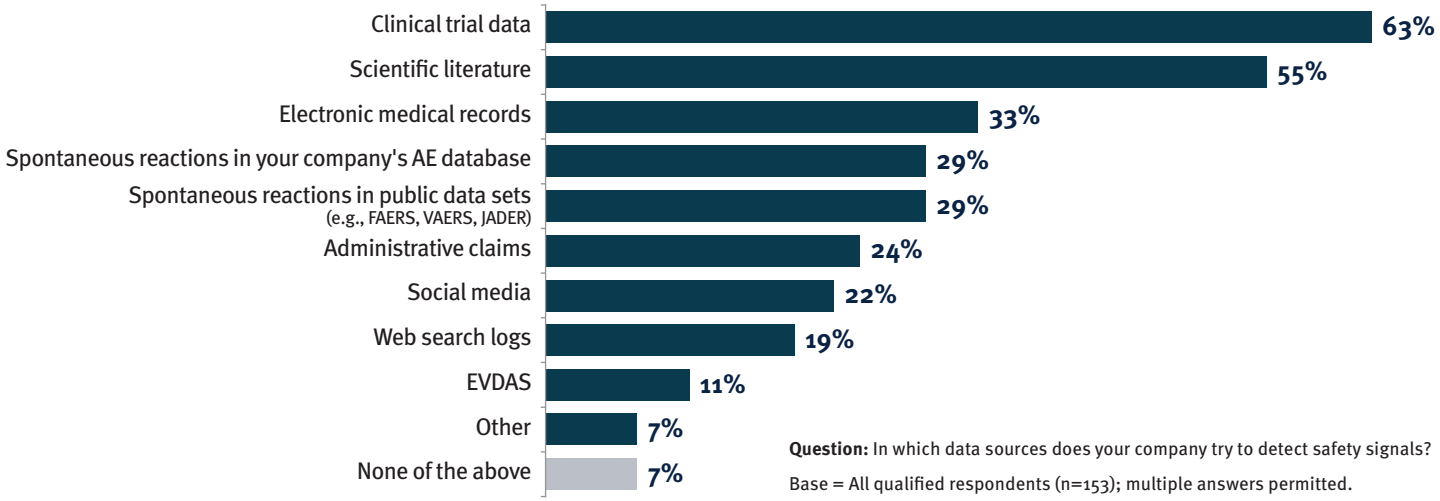
Even without considering the potentials for future possibilities, there are already tremendous pharmacoepidemiological possibilities associated with the cloud technology that are very difficult if not impossible to obtain otherwise. Insights into regional or temporal patterns of ADRs would be improved, the uncertainty about outcomes could be reduced, transparency would be significantly increased, and time lags between reporting and analysis could be decreased. Ultimately, this could result in more timely and qualitatively improved regulatory decision-making on important public health issues. And next to the direct implications on public health, we can also imagine how the integration of all the data obtained will allow for models to be built that are predictive for given individuals.

Cloud technology also brings big data applications in pharmacovigilance into play. While this is certainly a niche in PV today (despite the European CMDh task force on Big Data in Medicine), the pharmaceutical industry as well as the regulators will increasingly depend on big data solutions due to the increase in volume and variety of data (e.g., from social media, a wider use of patient-reported outcomes, data from wearables and other devices, etc.) to be analyzed in order to take the most informed decisions about the benefit-risk profiles of drugs (*See exhibit 4*). As we are aggregating more data, building models to generate knowledge out of the data, and applying the models on individual patients, we will also refine the models based on the comparison between real-world and modeled outcomes. Medicine will benefit from the significant maturity of information technology that will allow for entirely new approaches to improve our understanding of disease. We are certainly not going

Exhibit 4

Detecting Safety Signals – Data Sources

Respondents are most likely to try to detect safety signals in clinical trial data and scientific literature, followed at a distance by electronic medical records, spontaneous reactions in the company's AE database, and spontaneous reactions in public data sets.



Question: In which data sources does your company try to detect safety signals?

Base = All qualified respondents (n=153); multiple answers permitted.

to observe an overnight move of all of the data in medicine to be big data, but this move will be continuous and similar to what has been observed in other areas, and the evolution of new methods and applications will undoubtedly spur ever-new developments.

In very practical terms, and relevant for decision-making in pharmacovigilance, we need to have tools available to handle the massive volume and variety of data sources around adverse drug reactions. Ultimately, the output of big data after analysis by regulators and marketing authorization holders should be a set of consolidated, integrated data that are useful sources of knowledge about drugs and their desirable and undesirable events which can be utilized to prevent serious and severe adverse drug reactions at the level of the individual patient or at the public health level. And as regulators are going to use these methods, it would be quite advisable to marketing authorization holders as well in order to protect the position of their drugs in the market.

Finally, the storage of massive amounts of data in the cloud almost directly leads to the need to analyze data in a dislocated fashion there rather than going through tedious down- and uploading procedures. We are seeing many software solutions in the daily office environment being dislocated. More and more companies are moving away from licensing huge office software packages for their employees, packages that contain powerful tools that are only used (and actually required) by a minority of users, in favor of cloud-based solutions where rarely used parts are only bought ad hoc for a limited time and for limited use by the few users that actually require these capabilities. Why not utilize a similar approach for the analysis of drug safety data? Pharmacovigilance software providers offer highly specialized and very robust packages for the analysis of safety data. Companies, on the other hand, in particular smaller companies, require only a subset of these functions. The result is that many companies either use substandard or non-ideal solutions for the analysis of their data, or they forego software update cycles to save money. Given the importance of a thorough analysis of drug safety data and the potential waste of data that could give rise to new knowledge of drug biology, this is a deplorable state. Moving more of the pharmacovigilance data to be analyzed into the cloud should enable more users to explore, e.g. data mining and signal detection methodologies that have hitherto been unattainable for them.

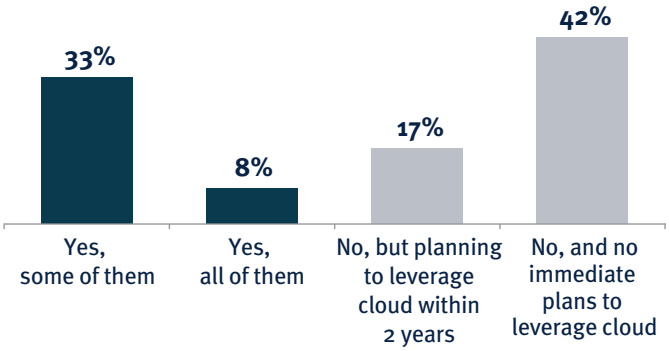
Therefore, it is not unexpected that almost 60% of respondents in the survey either have some or all of their safety solutions in the cloud or are planning to move there within the next two years (*See exhibit 5*). As with the general application of artificial intelligence solutions, a perceived hurdle to overcome for a wider application of cloud solutions is the associated cost. Similar to what we stated above, this is a rather interesting argument, as the experience from other industries shows the investment in cloud solutions to pay off very quickly. The significant number of survey respondents who are unsure about the value of PV cloud solutions undoubtedly contributes to the surprising concern about costs; education about the cloud's return on investment is needed to rectify this misconception. In addition, the prospect of always having access to the latest version of pharmacovigilance software without having to test and manage the installation in-house should be a convincing argument for a wider adoption

Exhibit 5

Safety Solutions In The Cloud

Four in ten respondents report at least some of their safety solutions are cloud-based. Those not currently leveraging cloud for safety solutions report data security concerns, data privacy concerns, cost and uncertain value/ROI as the most common deterrents.

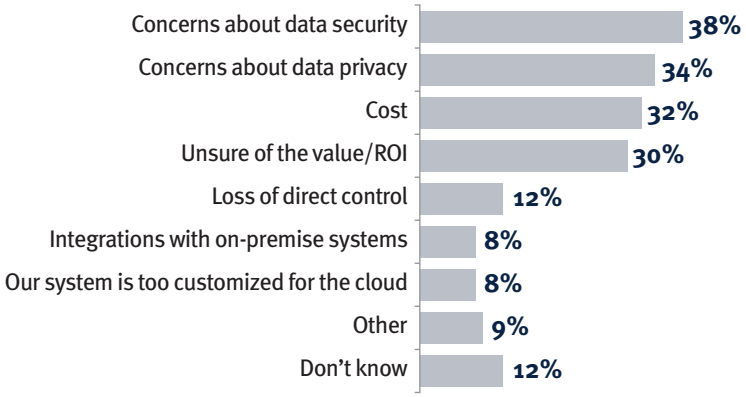
Are any of your safety solutions currently in the cloud?



Question: Are any of your safety solutions currently in the cloud?

Base = All qualified respondents (n=153).

Deterrents to Leveraging the Cloud

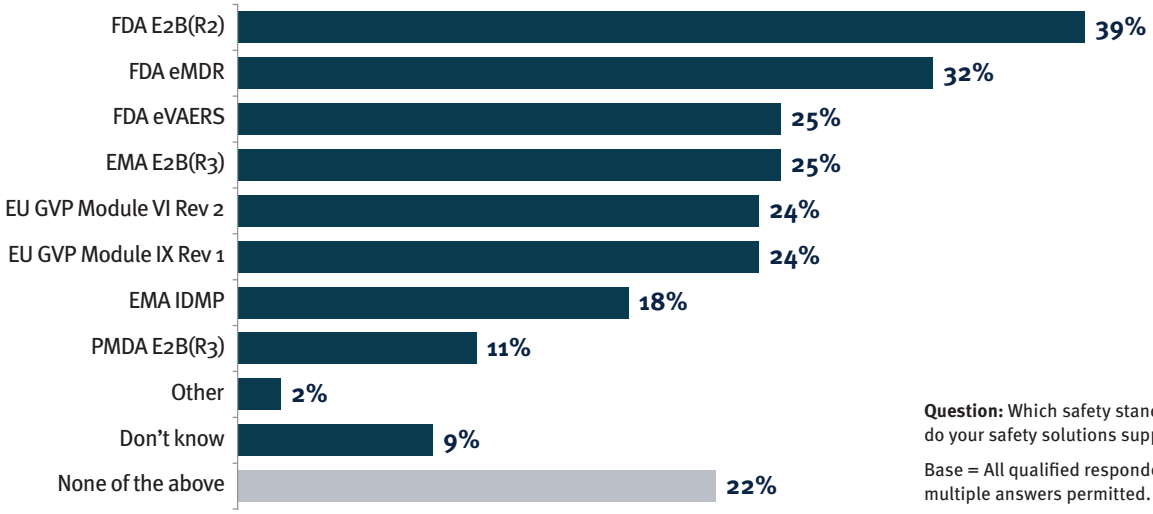


Question: What has deterred your company from leveraging the cloud for some or all of your safety solutions?

Base = Qualified respondents not currently leveraging cloud for safety solutions (n=74); multiple answers permitted.

Exhibit 6
Safety Solutions Standards/Guidelines Supported

Respondent company safety solutions are most likely to support FDA E2B(R2) and FDA eMDR.



Question: Which safety standards/guidelines do your safety solutions support today?

Base = All qualified respondents (n-153); multiple answers permitted.

of cloud-based software solutions in pharmacovigilance. Needless to say that software providers will have to meticulously demonstrate the validation of their solutions as required by different regulations, as well as compliance with all corresponding regulatory requirements, before pharmaceutical companies will be willing to accept more such solutions.

However, as the survey shows, the most significant deterrent to leveraging the cloud is a concern is about data security and privacy. Many countries are rightfully and severely concerned about data protection as we have observed massive breaches of data security all over the world, including having health-related data of thousands of citizens being compromised. This concern is obviously not new and not limited to health data, and mechanisms for encrypting, anonymizing, and otherwise securing information are being developed. As above, it will be important to show that any solution applied to patient data will adhere to the highest technological standards and will comply with all regulatory requirements. The development of these standards needs to be driven by industry together with regulators in order to obtain the trust from the public that is needed for the adoption of cloud solutions. It should not be forgotten either that data security is not necessarily higher in local-server-based databases than in the cloud; recent cyber-attacks on pharmaceutical companies' on-premise systems should make that clear. If the trust of patients in the security of their data is not given, the precautionary principle will take over and hinder technological advances that would ultimately benefit patients and public health. A balance will thus have to be found between the legitimate rights of citizens to control their data and the public health needs to quickly detect important safety signals and remove unsafe medicines from the market.

Regulatory requirements for reporting

Finally, we see that electronic solutions for the submission of cases are now used by almost all the respondents in the survey. While E2B(R2) and E2B(R3) as global gold standards are supported by the safety solutions of a majority of respondents, plus eMDR or eVAERS for devices and vaccines respectively, there are still over 20% of respondents that do not apply any of the common electronic solutions (*See exhibit 6*). Very likely this is due to an extremely low number of cases that does not warrant the investment into an electronic safety solution.

OVER **20%** of respondents have safety solutions that do not support modern regulatory standards and guidelines.

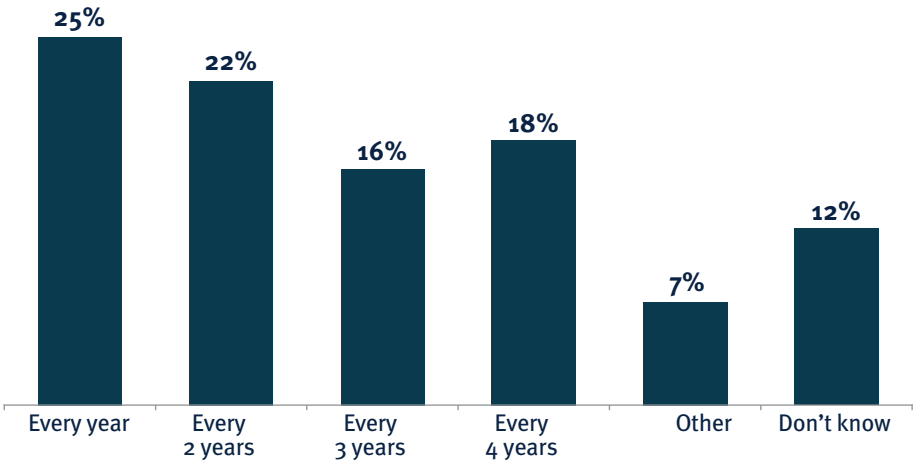
The new standards and high pace of change in the regulatory landscape in the last few years, together with different regional interpretations of reporting formats such as E2B(R3), have undoubtedly led to a fairly frequent upgrade schedule for PV applications, due to software vendors continuously providing new releases to be compliant with the latest laws. In the survey (*See exhibit 7*), almost half of respondents upgraded their systems at least every two years.

We have discussed the applications that make use of pharmacovigilance data above. All the knowledge generated can, however, be only as good as the data that feed into the analyses. The E2B(R3) report

Exhibit 7

Frequency Of Upgrading Safety Solutions (excluding patches)

Respondents upgrade their safety solutions at varied intervals. Just under half (47%) upgrade at least every two years.



Other responses:

- As growth dictates
- As needed
- CRO provided
- Depends of a lots of factors
- First development
- On an alert system-real time
- Only when rules changes or incident happens
- Varies
- When necessary

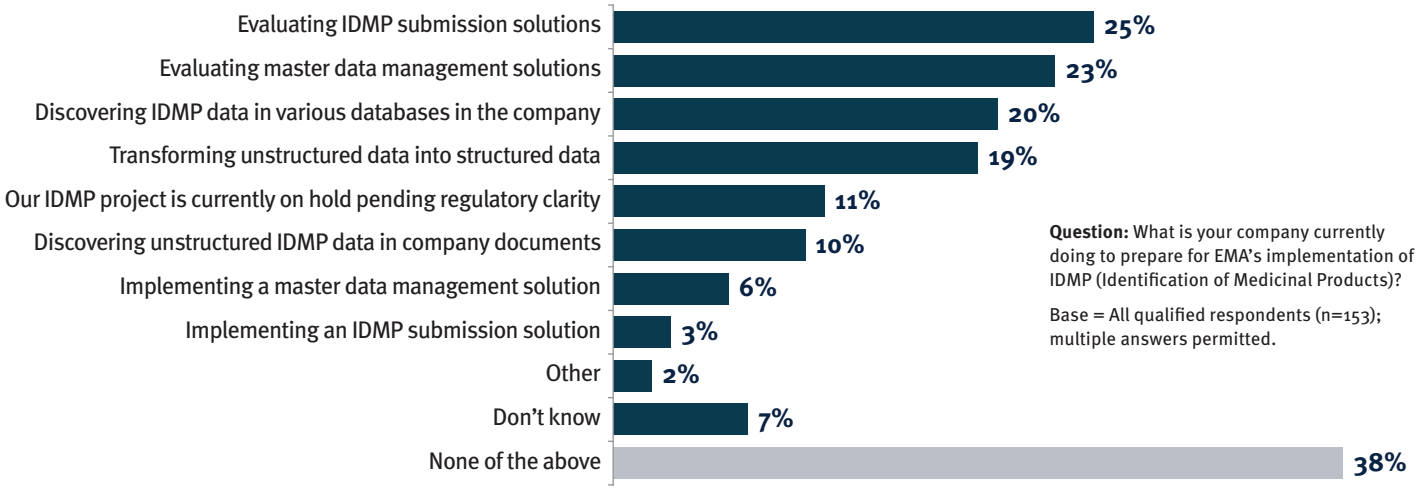
Question: On average and excluding patches, how often does your company upgrade your safety solutions?

Base = All qualified respondents (n=153).

Exhibit 8

Preparing For EMA's Implementation Of IDMP

Respondents lack unity with regard to their preparations for the EMA's implementation of IDMP. Approximately one in four are evaluating IDMP submission solutions and/or evaluating master data management solutions. One in five are discovering IDMP data in various databases in the company and/or transforming unstructured data into structured data. Just over a third are taking no such actions.



Question: What is your company currently doing to prepare for EMA's implementation of IDMP (Identification of Medicinal Products)?

Base = All qualified respondents (n=153); multiple answers permitted.

format has been developed with the use of data according to ISO IDMP standards. If we ultimately want to make good use of the data collected in supranational regulatory safety databases like EudraVigilance or in our in-house company databases (whether or not they are in the cloud or on local servers), these data need to be structured appropriately and according to modern standards.

As even small regulatory agencies are increasingly moving away from paper-based, fax or e-mail reporting, there may be a need for simple solutions for smaller companies with a limited number of reportable cases in the future in order for them to remain compliant and – most importantly – to enable their data to be integrated as high-quality data into the global safety databases required for effective and efficient pharmacovigilance. Again, such solutions could be cloud-based to allow for a resource- and cost-conscious implementation of

the latest technological standards. For all companies, as well as for regulatory agencies, a swift completion of the implementation of IDMP standards, which are meant to ensure wide interoperability across global regulatory and healthcare communities, is required. This is critical in ensuring accurate analysis of safety data and unambiguous communication between stakeholders.

Therefore the rather low readiness for IDMP implementation as exemplified in the survey is remarkable (*See exhibit 8*) and should urgently be tackled by companies as well as regulators, if we want to have suitable data for the innovative data management and analysis solutions being developed.

Conclusion

We have looked at three important developments around drug safety data and their analysis and how industry is prepared for them. While big data and cloud solutions are being utilized by many companies to some extent (with some completely embracing them), there remain concerns about data security and integrity that need to be addressed by all stakeholders. The data that feed into safety databases need to be of good quality and appropriately structured in order to be meaningfully analyzed. Not all respondents in the survey are prepared for the latest developments and regulatory requirements in data standards and obviously have a need to urgently catch up. The potential to use artificial intelligence methods increasingly for the analysis of the increasing amounts of pharmacovigilance data is well understood and many companies are moving (or planning to move) there, and we can predict that routine tasks in pharmacovigilance will in the future be increasingly automated. It will be crucial, however, for regulatory authorities to very clearly provide a position about the use of AI as well as the acceptable level of quality from AI applications. But in parallel with the shaping of those definitions, given the massive increase in their AE case workloads that most companies are currently experiencing, the industry will out of necessity proceed swiftly with the adoption of AI and cloud technologies to reduce their costs and increase their efficiencies.

Like other industries, the pharmaceutical business and in particular the pharmacovigilance field will see a massive change in their processes in the near future, away from tedious, repetitive manual tasks towards a better utilization of scarce resources, in particular medical and scientific knowledge, for value-adding tasks. It is imperative for all stakeholders – industry, service providers and regulators – to provide an environment in which such a transformation can take place without ever compromising public health or the safety of the individual patient, and ideally providing additional benefit for patients.