ICMRA Pharmacovigilance

A Draft Policy Paper for ICMRA Pharmacovigilance Project

ICMRA Big Data Working Group
(Australia, Brazil, Canada, European Medicines Agency, Italy, Japan, the Netherlands, New Zealand, Singapore, South Africa, Sweden, Switzerland and the United Kingdom)

Version 2.3.1

Sept 2016

Executive Summary

The ICMRA pharmacovigilance working group was formed to examine the gaps in pharmacovigilance. A sub-group of this working group, including Australia (Therapeutic Goods Administration [TGA]), Brazil (Agência Nacional de Vigilância Sanitária [ANVISA]), Canada (Health Canada), the European Medicines Agency (EMA), Italy (Agenzia Italiana del Farmaco [AIFA]), Japan (Pharmaceuticals and Medical Devices Agency [PMDA]), the Netherlands (Medicines Evaluation Board [MEB]), New Zealand (Ministry of Health [MOH]), Singapore (Health Sciences Authority [HSA]), South Africa (Medicines Control Council [MCC]), Sweden (Medical Products Agency [MPA]), Switzerland (Swissmedic) and the United Kingdom (Medicines and Healthcare Products Regulatory Agency [MHRA]) are working together to address opportunities and challenges for the use of "big data" in pharmacovigilance. A contribution was also received from Ireland (Health Products Regulatory Authority [HPRA]).

This paper has been prepared to stimulate discussion among the ICMRA pharmacovigilance sub-group members, to develop a draft policy paper on the opportunities and challenges for big data and analytics within the context of pharmacovigilance, and to facilitate international collaboration. The development of this paper was coordinated by Health Canada with contributions from each member. It represents a snapshot of the experiences within each member agency at September 2016 in the rapidly evolving frame of big data and analytics for pharmacovigilance.

The spontaneous reporting systems (SRSs) described by sub-group members contain structured data in quantities that could reasonably be processed on a single machine. These systems alone do not meet the definition of "Big Data". All sub-group members collect domestic spontaneous reports in which suspected adverse drug reactions (ADRs) are reported by health care professionals and companies, and stored in a database system.

All sub-group members are in agreement that while SRSs are crucial and essential components of pharmacovigilance, they can never give a complete picture of patient safety information. Some of the limitations of SRSs are:

- The number of reports received cannot be used as a basis for determining the incidence of a reaction as neither the total number of reactions occurring in the population nor the number of patients exposed to a health product is known.
- The data collected often have limited patient information including medical histories, concomitant treatment(s), pre-existing conditions, time to onset, etc.
- There is under-reporting of adverse reactions with both voluntary and mandatory surveillance systems, and reporting rates may vary widely for drugs as well as for jurisdictions.
- Numerical comparisons cannot be made between reactions associated with different health products on the basis of the data in the line-listings.

- The data reported do not represent all known safety information concerning the suspected health product(s) and cannot be used in isolation to make decisions regarding an individual's treatment regimen; other sources of information, including the prescribing information for the product, should be consulted.

These limitations brought this sub-working group to find new ways to enhance the quality and quantity of ADRs, including the use of "big data" in pharmacovigilance.

Big data refers to a collection of structured and unstructured data. A definition of big data in post-market surveillance is lacking at this time. In post-market surveillance, structured data may include SRSs (with individual case safety reports [ICSRs]), electronic health records (EHRs), electronic medical records (EMRs), administrative health data (AHD), registries, etc. Unstructured data may include social media like Twitter, Facebook, patient forums, clinical narratives within EMRs, etc.

Sub-group members are interested in evaluating big data analytics for pharmacovigilance because it has the potential to supplement traditional spontaneous reporting systems in several ways:

- It could be a cost-effective (re)use of existing administrative data to enable active surveillance.
- It could provide more timely signal detection over traditional spontaneous reporting systems.
- It has the potential to complement traditional systems and use epidemiological methods to estimate the incidence of adverse events in populations.
- It could provide a better ability to identify and investigate medicine-adverse event associations that happen over a longer period of time (which could be missed by spontaneous reporting).
- It enhances the ability to investigate signals across different sub-populations and to control for confounders.

All members agreed to share more experiences and knowledge in utilizing big data for regulatory purpose as a first step. Such activity can further identify common challenges and gaps for regulatory harmonization and collaboration on use of big data in pharmacovigilance. The experiences of the sub-group members are summarized in the tables below and detailed in Section 6, 7, and 8. Sub-group members are at various stages of evaluation, development and validation of big data sources to supplement traditional spontaneous reporting. Levels of maturity range from expressing an interest in evaluating the data source to routine use of a data source to validate signals detected from spontaneous reporting.

Table 1. Big Data Sources of Interest for Pharmacovigilance (PV)

Country	Data Source	Data Source Type	Stage of Use for PV
Australia	1. Medicare Benefits Schedule (MBS) 2. Pharmaceutical Benefit Schedule (PBS) 3. Admitted Patients Data Collection 4. My Health Record (MHR) 5. Australia Childhood Immunisation Register (ACIR) 6. Medicine Insight	1. Administrative Health data 2. Administrative Health data 3. Clinical/Administrative Health Data 4. Electronic Health Record 5. Clinical register 6. Electronic Medical Records from approximately 500 General Practices	1. Proof-of-Concept 2. Proof-of-Concept 3. Interest 4. Interest 5. Interest 6. Interest
Brazil	1. ?	1. ?	1. ?
Canada	1. Canadian Institute for Health Information (CIHI) 2. Canadian Primary Care Sentinel Surveillance Network (CPCSSN) 3. IMS Evidence 360 4. Drug Safety and Effectiveness Network (DSEN) Social Media Study 5. Canada Vigilance	1. Administrative data 2. Distributed Electronic Medical Records 3. Electronic Medical Records 4. Social Media 5. Spontaneous ADRs	1. Interest 2. Proof-of-Concept 3. Interest 4. Interest 5. Routine
Ireland	1. None currently available	1. N/A	1. N/A
Italy	1. National Pharmacovigilance Database 2. AIFA National Registers	1. Spontaneous ADRs 2. Electronic Medical Records from Prescribers	1. Routine
Japan	1. Medical Information Database-Network (MID-NET) 1. Medical Information for Risk Assessment Initiative (MIHARI)	1. Distributed Electronic Medical Records 1. Multiple types of databases including claims database	1. Pilot and Validation 1. Routine since 2014
Netherlands	1. Lareb 2. Pharmo 3. IPCI 4. ERGO 5. LASA 6. NIVEL Primary Care database 7. IKNL 8. WEB-RADR	1. ADR reports from HCPs and patients 2. Electronic Medical Records registries, linked to registries (eg cancer, pathology, perinatal) 3. Electronic Medical Records from GPs 4. Population-based cohort study in elderly living in Rotterdam 5. Population-based cohort study in elderly in multiple areas in NL 6. Electronic Medical Records from GPs 7. Dutch cancer registry 8. Collaboration to Mine	1. Routine 2. Research 3. Research 4. Research 5. Research 6. Routine and Research 7. Research

		Social Media for PhV	
New Zealand	1. ?	1. ?	1. ?
Singapore	1. National Electronic Health Record	1. Electronic Health Record	1. Interest
South Africa	1. Pregnancy Registry 2. Targeted Spontaneous reports 3. Spontaneous reports 4. Surveillance and health information system 5. Published or unpublished local studies	1. Sentinel cohort 2. ADR reports 3. ADR reports 4. Administrative health data, electronic laboratory, admission and pharmacy records 5. HIV/AIDS and TB Cohorts hospital active ADR surveillance studies, and clinical trials	1. Routine 2. Routine 3. Routine 4. Under exploration 5. Routine
Sweden	1. Quality Registries 2. Medical Record System 3. National Pharmacovigilance Database	1. Spontaneous ADRs 2. Distributed Electronic Medical Records 3. Spontaneous ADRs	1. Routine and Interest 2. Projects ongoing 3. Routine
Switzerland	1. SRS	1. Spontaneous ADRs	1. Interest
UK	1. Clinical Practice Research Datalink (CPRD) 2. WEB-RADR	1. Distributed Electronic Medical Records 2. Collaboration to Mine Social Media for PV	1. Routine 2. Research

Table 2. Summary Data Sources Volume and Coding

Country	SRS		EMR/EHR data	Administrative Health data
	Volume	Coding	Coding	Coding
Australia	380,000 domestic reports (1971-April 2016)	ICH-E2B MedDRA	SNOMED-CT-AU MBS/PBS-specific coding AMT	MBS/PBS-specific coding
Brazil	225,078 domestic reports? (2007-2014)	?	?	?
Canada	500,000 domestic reports (1967-2015)	ICH-E2B MedDRA	ICD-10-CA	ICD-10-CA ATC
Ireland	61,021	ICH-E2B MedDRA	ATC	ICD-10 ATC
Italy	372,672 domestic reports (literature included) (1990-2016)	ICH-E2B MedDRA	N/A	ICD-10
Japan	390,669 domestic reports (2004Q2-2015Q3)	ICH-E2B MedDRA	ICD-10-codes	ICD-10-codes
NL: Lareb	96.000 [2010-2015]	ICH-E2B MedDRA	ATC/ICD-10	ICPC/ICD-10/ATC
New Zealand	96,000 domestic reports (1965-April 2016)	WHO-ART MedDRA	ICD-10 ATC	ICD-10
Singapore	195,300 domestic reports (1993- April 2016)	ICH-E2B (in progress) WHO-ART, bridge to	ICD-10 SNOMED-CT	ICD-10

		MedDRA		
South Africa	31,585 (domestic reports) (1992-Aug 2016)	WHO-ART	ICD-10	N/A
Sweden	150,000 domestic reports (1965-2015)	ICH-E2B MedDRA	ICD 10	NA
Switzerland	83,000 domestic reports	ICH-E2B MedDRA	tbd	tbd
UK	800,000 domestic reports (1971-April 2016)	ICH-E2B MedDRA	SNOMED-CT	NA

Table 3. Data Mining Survey

Country	**Data mining**			
	SRS	**Healthcare system**		**Non-conventional**
		EMR	**AMR**	**Social media**
Australia	Yes	No	Yes MBS PBS	No
Brazil	No	No	No	No
Canada	No	No	No	No
Ireland	No	No	No	No
Italy	No	No	No	No
Japan	Yes	Yes MID-NET MIHARI	Yes MIHARI	No
Netherlands	Yes: Lareb	?	?	Yes: WEB-RADR since 2015
New Zealand	Yes	No?	No?	No
Singapore	Yes	Pilot	No	No
South Africa	Yes	No	Under exploration	No
Sweden	Yes	National Database	No	No
Switzerland	No	No	No	No
UK	Yes	Yes CPRD	?	Yes: WEB-RADR in 2015

While each member acknowledges benefits of big data in post-market surveillance, a number of challenges were highlighted, particularly with the integration and utilization of big data into the current pharmacovigilance framework. The common challenges identified include:

Security and privacy

- Privacy & legal considerations for sharing of data between countries

Partnerships

- Building domestic and international partnerships to access data sources that are not held by the regulator.

Infrastructure & capacity

- Building internal capacity to leverage new data sources, including both IT infrastructure and the expertise to use it effectively.
- Building infrastructure and common data models to support datasets and develop innovative methods to link together different data sources.
- Resources required to develop, evaluate and use data may be substantial and will involve multi-disciplinary teams.

Standards

- Developing international standards, principles and best practices for sharing, combining and validating data from multiple sources.
- Establishing international harmonized standard to accept results of big data analysis by other regulatory agencies.
- Developing common data models to share data.
- Different coding systems for spontaneous AE databases and health care data, and requires extraction and mapping of ADRs.

Data analysis

- Establishing methods and standards to evaluate data for secondary uses, which may be biased due to a number of factors, including practitioner behaviour, administrative dataset is mainly for financial purposes, EMR is primary for clinical management.
- Incomplete data due to partial participation by healthcare providers (public vs. private).
- Challenges in converting free text data to structured data.
- The rules governing data collection may change over time, and electronic access to historical records may be limited.

Despite of all the common challenges, all sub-group members have expressed an interest in collaborating further on big data initiatives for pharmacovigilance. However, a more focussed work- sharing is required for the benefit of this sub-working group because of the following:

- All sub-group members are already participating in the World Health Organization-Uppsala Monitoring Centre (WHO UMC) i.e., all are submitting Individual Case Safety Reports (ICSRs) to the VigiBase. Therefore, sharing ICSRs within sub-group members will be duplication of work since these data are already available in the VigiBase.
- For health care data, data sharing will be hindered by privacy and other laws within and across each member jurisdictions. In addition, other methods are required to assist with the unique issues that arise from working with non–ICSR data.
- It is noted that other ICMRA members in addition to those that contributed to this paper, including the US FDA and the EMA, have various pharmacovigilance big data initiatives at various stages of maturity. Current regulatory initiatives and experiences regarding validation studies for the use of real-world data in pharmacovigilance of all ICMRA

members could be shared, to minimize redundancy, and the best practices and validation findings of these initiatives could be shared when they become available.

- The integration and utilization of non-conventional/unstructured data sources in pharmacovigilance is also resource-intensive and may be potentially redundant. An established framework is required for knowledge transfer of the results of existing research initiatives to validate these data sources for use in pharmacovigilance. For example, issues surrounding such data can be discussed among all sub-group members based on analytical results shared by an agency.

Based on the sub-group members' contributions in Section 6, 7, and 8, a two-stage approach for working sharing maybe considered.

- **Short-term:** All sub-group members have expressed an interest in a more focussed knowledge sharing on a technical level to identify specific challenges and gaps. In particular, sub-group members raised the needs of best practices for utilizing real-world data sources such as EHRs, EMRs and AHD together with traditional SRS data, either in a single database or in a distributed method. Ensuring that aggregated data from each member are consistent and comparable is a prerequisite to more focussed research collaborations. The development of guidance or technical standards, if this is feasible, would be beneficial to all sub-group members.
- **Long-term:** Collaborative development of study protocols, including a common observational study protocol on a project of common interests for each member to conduct within their databases; and a meta-analysis study protocol with common data extraction procedure for each member's databases. These approaches will facilitate collaboration and will overcome the privacy and security issues.

Based on the two-stage approach above, recommendations were prepared in order, with earlier recommendations understood to be necessary prerequisites to those that follow. The following recommendations were made from the sub-group members to the ICMRA Management Committee to promote additional collaboration:

- A standing working group should be formed to facilitate knowledge transfer between ICMRA members with respect to leveraging big data analytics for pharmacovigilance.
- All ICMRA members should be invited to notify partners of on-going research initiatives and validation studies in using non-conventional data sources to complement existing pharmacovigilance activities, to minimize redundancy.
- All ICMRA members should be invited to share the results of research initiatives and validation studies in using big data sources to complement existing pharmacovigilance activities, when they become available.
- Best practices, including coding standards, for combining multiple sources of real-world data, including EHRs, EMRs and AHD with traditional SRS data, should be developed and shared.

- Develop a common observational study protocol on a project of common interests for each member to conduct within their databases. Findings are then analyzed and shared within members.
- Develop a meta-analysis study protocol with common data extraction procedure for each member's databases.

	Names	Organization
Prepared by	Daniel Buijs	HC
	Jhona Rose	HC
	Amy Taylor	HC
	Vikesh Srivastava	HC
	Duc Vu	HC
Contributors		
	AIFA	AIFA
	Livia Ramalho Evangelista	ANVISA
	Joey Gouws	MCC
	Alison Cave	EMA
	Peter Arlett	EMA
	Benjamin Ong	HSA
	Ang Pei San	HSA
	Belinda Foo	HSA
	Niamh Buckley	HPRA
	Mick Foy	MHRA
	Jonathan Mogford	MHRA
	Birte van Elk	MEB
	Chris James	MOH
	Kerstin Jansson	MPA
	Yoshiaki Uyama	PMDA
	Naoyuki Yasuda	PMDA
	Chieko Ishiguro	PMDA
	Valeriu Toma	Swissmedic
	Daniel B. Huerlimann	Swissmedic
	Bronwen Harvey	TGA
	Lara Hooke	TGA
	Richard Hill	TGA
	Margaret Wilson	TGA
	Sophie Sommerer	HC
	Samina Aziz	HC
	Geeta Saxena	HC
	Basanti Ghosh	HC
	Vicky Hogan	HC

Table of Contents

Abbreviations, Acronyms, and Definitions

Abbreviation and Acronyms

ACIR	Australian Childhood Immunisation Register
AIR	Australian Immunisation Register
ADR	Adverse Drug Reaction
AE	Adverse Event
AEFI	Adverse Events Following Immunization
AHD	Administrative Health Data
AMT	Australian Medicines Terminology
AIFA	Agenzia Italiana del Farmaco
ANVISA	Agência Nacional de Vigilância Sanitária
ATC	Anatomical Therapeutic Chemical Classification System
BCPNN	Bayesian Confidence Propagation Neural Network
CARM	Centre for Adverse Reactions Monitoring
CDM	Common Data Model
CHM	Commission for Human Medicines
CIHI	Canadian Institute for Health Information
CPRD	Clinical Practice Research Datalink
CPCSSN	Canadian Primary Care Sentinel Surveillance Network
DHCPL	Dear HealthCare Professional Letters
DSEN	Drug Safety and Effectiveness Network
EHR	Electronic Health Records
EMA	European Medicines Agency
EMR	Electronic Medical Records
FDA	Food and Drug Administration

GPS	Gamma Poisson Shrinker
HC	Health Canada
HSA	Health Sciences Authority
IC	Information Component
ICMRA	International Coalition of Medicines Regulatory Authorities
ICSR	Individual Case Safety Report
IMI	Innovative Medicines Initiative
JADER	Japanese Adverse Drug Event Report database
MAH	Market Authorization Holder
MBS	Medicare Benefits Schedule
MedDRA	Medical Dictionary for Regulatory Activities
MERP	Medication Error Reporting Programme
MGPS	Multi-item Gamma Poisson Shrinker
MHR	My Health Record
MHRA	Medicines and Healthcare Products Regulatory Agency
MIHARI	Medical Information for Risk Assessment Initiative
MID-NET	Medical Information Database-Network
MOH	Ministry of Health
MOHH	Ministry of Health Holdings
MPA	Medical Products Agency
MRAs	Medicines Regulatory Authorities
MWS	Medical Warning System
NADEMC	National Adverse Drug Events Monitoring Centre
NEHR	National Electronic Health Record
NHI	National Health Index

NHS	National Health System
NSW	New South Wales
NRIC	National Registration Identity Card
NZPhVC	New Zealand Pharmacovigilance Centre
OMOP	Observational Medical Outcomes Partnership
PBS	Pharmaceutical Benefit Schedule
PMDA	Pharmaceuticals and Medical Devices Agency
PRR	Proportional Reporting Ratio
ROR	Reporting odds ratios
SNOME-CT	Systemized Nomenclature of Medicine Clinical Terms
SRS	Spontaneous Reporting System
TGA	Therapeutic Goods Administration
UK	United Kingdom
WHO	World Health Organization
WHO-ART	World Health Organization-Adverse Reaction Terminology
WHO-UMC	World Health Organization-Uppsala Monitoring Centre

Definitions

Big data: Refers to collections of structured and unstructured data that may be enormous, in the range of several billion gigabytes.

Pharmacovigilance: WHO refers to as "the science and activities relating to the detection, assessment, understanding and prevention of adverse effects and all other problems related to medicines."

Signal: WHO refers to as "Reported information on a possible causal relationship between an adverse event and a drug, the relationship being unknown or incompletely documented previously."

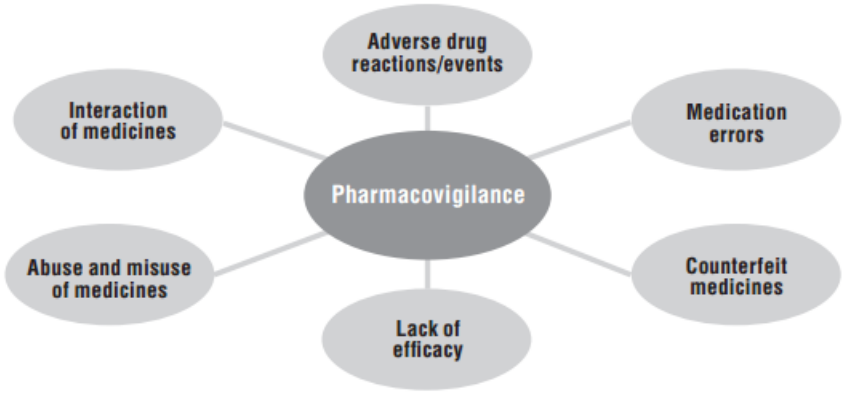
Signal detection: Refers to the "process that aims to find, as soon as possible, any indication of an unexpected drug safety problem which may be either new ADRs or a change of the frequency of ADRs that are already known to be associated with the drugs involved."

Structured: In post-market surveillance, structured data may include spontaneous reporting systems, electronic health records (EHRs), electronic medical records (EMRs), administrative health data (AHD), registries, etc.

Unstructured: May include social media like Twitter, Facebook, patient forum, clinical narratives within EMR, etc.

1 Introduction

The World Health Organization (WHO) defines Pharmacovigilance as "the science and activities relating to the detection, assessment, understanding and prevention of adverse effects and all other problems related to medicines."[1] The scope of pharmacovigilance is complex and includes adverse drug reactions (ADRs) or events, medication errors, counterfeit or substandard medicines, lack of efficacy, misuse and/or abuse, and interaction between medicines (see diagram).[2]



Pharmacovigilance relies heavily on spontaneous reporting in which suspected adverse drug reactions (ADRs) are reported by health care professionals, manufacturers or directly by patients. Spontaneous reporting systems (SRSs) provide the highest volume of information on patient safety related either to the products themselves or to their use. The most important function of SRSs is early detection of signals.[4]

A signal is defined by the WHO as "Reported information on a possible causal relationship between an adverse event and a drug, the relationship being unknown or incompletely documented previously."[3] Signal detection refers to the "process that aims to find, as soon as possible, any indication of an unexpected drug safety problem which may be either new ADRs or a change of the frequency of ADRs that are already known to be associated with the drugs involved."[4]

While SRSs are crucial and essential components of pharmacovigilance, they can never give a complete picture of patient safety information. Some of the limitations are:[5]

[1] http://apps.who.int/medicinedocs/pdf/s4893e/s4893e.pdf
[2] http://www.who.int/medicines/areas/quality_safety/safety_efficacy/EMP_PV_Indicators_web_ready_v2.pdf
[3] http://www.who-umc.org/DynPage.aspx?id=115092&mn1=7347&mn2=7252&mn3=7613&mn4=7614
[4] http://www.dsru.org/consulting-on-risk-management/signal-detection/
[5] http://www.hc-sc.gc.ca/dhp-mps/medeff/databasdon/conditions_search-recherche-eng.php

- The number of reports received cannot be used as a basis for determining the incidence of a reaction as neither the total number of reactions occurring in the population nor the number of patients exposed to a health product is known.
- The data collected often have limited patient information including medical histories, concomitant treatment(s), pre-existing conditions, time to onset, etc.
- There is under-reporting of adverse reactions with both voluntary and mandatory surveillance systems, and reporting rates may vary widely for drugs as well as for jurisdictions.
- Numerical comparisons cannot be made between reactions associated with different health products on the basis of the data in the line-listings.
- The data reported do not represent all known safety information concerning the suspected health product(s) and cannot be used in isolation to make decisions regarding an individual's treatment regimen; other sources of information, including the prescribing information for the product, should be consulted.

These limitations force all regulatory agencies to find new ways to enhance the quality and quantity of ADRs, including the use of "big data" in pharmacovigilance.

The International Coalition of Medicines Regulatory Authorities (ICMRA) is a voluntary, executive level, strategic coordination, advocacy, and leadership entity of national and regional medicines regulatory authorities (MRAs) that work together to provide direction for a range of areas and activities common to many regulatory authorities' missions and goals.[6] ICMRA pharmacovigilance working group was formed to examine the gaps in post-market pharmacovigilance.

A subgroup of this working group including Australia (Therapeutic Goods Administration [TGA]), Canada (Health Canada), Japan (Pharmaceuticals and Medical Devices Agency [PMDA]), Singapore (Health Sciences Authority [HSA]), and the United Kingdom (Medicines and Healthcare Products Regulatory Agency [MHRA]) was formed to address opportunities and challenges for "big data" in the pharmacovigilance framework. Due to significant interest, the sub-group was later expanded to include Brazil (Agência Nacional de Vigilância Sanitária [ANVISA]), the European Medicines Agency (EMA), Italy (Agenxia Italiana del Farmaco [AIFA]), the Netherlands (Medicines Evaluation Board [MEB]), New Zealand (Ministry of Health [MOH]), South Africa (Medicines Control Council [MCC]), Sweden (Medical Products Agency [MPA]), and Switzerland (Swissmedic). A contribution was also received from Ireland (Health Products Regulatory Authority [HPRA]). Health Canada is the lead and will prepare and collate all information provided from the other sub-group members.

[6] https://www.tga.gov.au/international-coalition-medicines-regulatory-authorities-icmra

2 Purpose

This paper is prepared to stimulate discussion among the ICMRA pharmacovigilance sub-group members, to develop a draft policy paper on the opportunities and challenges for big data and analytics within the context of pharmacovigilance, and to facilitate international collaboration.

3 Scope

The development of this paper was coordinated by Health Canada with contributions from HSA, MHRA, PMDA, and TGA. It reflects recent thinking within each member agency on the experiences on big data and analytics in pharmacovigilance.

4 What is "Big Data and Analytics"?

4.1 "Big Data" in pharmacovigilance

Big data refers to a collection of structured and unstructured data that may be enormous, in the range of several billion gigabytes. In post-market surveillance, structured data may include SRSs, electronic health records (EHRs), electronic medical records (EMRs), administrative health data (AHD), registries, etc. Unstructured data may include social media like Twitter, Facebook, patient forums, clinical narratives within EMRs, etc. A definition of big data in post-market surveillance is lacking at this time although the EMA refers to big data as "an umbrella term describing large data sets from any source." [7]

Other regulatory agencies, like the United States Food and Drug Administration (US FDA)[8] and European Medicines Agency (EMA)[9] have already incorporated AHD and EMRs as part of their post-market surveillance systems. Further research is required to demonstrate the benefit of using unstructured data in post-market surveillance.

While benefits of big data in post-market surveillance are increasingly being recognized, its integration and utilization to the current post-market surveillance frameworks requires clearly defined objectives and plans, and financial resources. Big data is characterised by 4Vs; these are volume, variety, velocity and veracity. Each characteristic requires a large amount of financial planning and human resources for the development, implementation, validation and maintenance. [10,11] Some of these are summarize below:

- The **volume** of data combined from different sources (structured and unstructured) may be enormous, in the range of several billion gigabytes. This requires extensive data storage capacity and data processing power.

[7] STAMP Commission expert group meeting on March 10, 2016
[8] http://www.fda.gov/downloads/ScienceResearch/DataMiningatFDA/UCM443675.pdf
[9] http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2009/11/WC500011434.pdf
[10] http://bok.ahima.org/doc?oid=105683#.VyO2Jvkwjq4
[11] http://www.ibmbigdatahub.com/infographic/four-vs-big-data

- The **variety** of data, often with hidden differences. Each dataset is unique because of the purpose, design, and processes of how information are collected and stored differ. The varied nature of datasets may make it difficult to combine data from different datasets with the same purposes (i.e. EMRs), and requires linking information to combine data from different datasets (i.e. EMRs and AHD). When linking data from multiple sources that only partially capture a population of interest, there is potential for both significant gaps in coverage or unrecognized redundancies to lead to bias.
- The **velocity** at which the data is processed and analyzed. The time it takes to enter data in the system affects the time it will take to process and analyze data. This is interconnected with how the data are stored, transformed, linked and extracted. Each step has an important role in providing timely results. Data mining can speed up the process of signal detection, but the limiting step is the time required for data to accrue. When considering collections of data from multiple sources, the rate at which new information accumulates may exceed the capacity of a single system to record or collect information in an economically feasible way. Fully-automated and semi-automated strategies to prioritize future data collection based on interim analyses have been developed, which help balance the resources required to perform analyses on big data.
- The **veracity** of data, because of errors, biases, noise and uncertainty affect the validity of results. The detection and interpretation of signals depend on the quality of data stored in the database. Manual data preparation and entry is not only time consuming, but is also prone to human error. Automated systems improve the accuracy, but are not always feasible and poorly designed systems may introduce bias. Differences in coding conventions, mapping of data elements from each system, missing information, and duplicates may create further errors when linking databases.

4.2 Analytics

The current technique for signal detection in spontaneous reporting databases is called data mining. Data mining methods use statistical tools to examine large datasets to identify drug-associated adverse events (signals) that are reported more frequently than expected. This is done by estimating the expected reporting frequencies on the basis of information on all drugs and all events in the database. It is assumed that, in the absence of disproportionality, the distributions of reported adverse events are the same across drugs. If a specific adverse event is associated with a given drug, this event would have higher reporting frequency and create reporting disproportionality. It is important to note that an automated signal is not necessarily a true signal (adverse drug effect). It is exploratory and hypothesis generating, and is subjected to further investigation in order to qualify if the automated signal is a credible adverse drug effect.[12]

[12] http://fusion.isif.org/proceedings/Fusion_2011/data/papers/200.pdf

Several algorithms have been developed and implemented for disproportionality analysis. All of these algorithms calculate signal scores to assess whether there is a statistically significant association between a drug and an adverse event. These include the following:

- Relative Reporting Ratio (RRR)
- Proportional Reporting Ratio (PRR)
- Reporting Odds Ratio (ROR)
- Gamma Poisson Shrinker (GPS)
- Multi-Item Gamma Poisson Shrinker (MPGS)
- Information Component (IC) derived from Bayesian Confidence Propagation Neural Network (BCPNN) analysis.

Although these methods are highly applicable in spontaneous reporting systems, they are potentially applicable for big data as well. The larger the dataset, the greater the statistical power to detect the disproportionality. The effectiveness of all data mining methods highly depends on the size and quality of the data. The following should be considered:[13]

- Advantages and disadvantages of each method;
- Sensitivity and specificity of each method; and
- Threshold for each method

A balance between sensitivity (true signal), specificity (true non-signal), false positive (not a true signal) and false negative (not a true non-signal) is required in order to establish an optimal ratio of signal and background noise. False negative signals are potentially harmful because important events may have been missed. In contrast, a large number of false positive signals is also a problem because they can be costly and time-consuming to validate. In addition, false positive signals can be costly as well if decisions with public health impact are based on faulty evidence. Therefore, thresholds for sensitivity and specificity must be established and defined precisely.[12]

For example, the PRR is a very sensitive method that may generate a high number of false positive signals, particularly when the number of reports is small because it does not adjust for small observed or expected numbers of reports of the product event pair of interest. Therefore, case count thresholds of number of reports > 3 are also used in association with the PRR and Chi-square statistics to reduce the number of false positives. The threshold of at least 3 reports may vary according to the extent of usage of health products and the potential public health impact.[9]

Each method will provide an estimate of disproportionality according to its inherent assumptions. For example, BCPNN thresholds rely on the IC calculation and the 95%

[13] http://cdn.intechopen.com/pdfs/38579/InTech-Data_mining_techniques_in_pharmacovigilance_analysis_of_the_publicly_accessible_fda_adverse_event_reporting_system_aers_.pdf

confidence interval as criteria for its signal detection methodology, whereas the MGPS uses the Poisson distribution to model the observed count for each product-event combination.

5 Relevant Data Sources

5.1 Spontaneous Reporting

SRSs are designed to gather individual case safety reports (ICSRs) of suspected ADR from a variety of sources, including clinicians, pharmacists, other healthcare professionals, pharmaceutical companies, medical literature, patients and the general public. They remain the cornerstone of pharmacovigilance; they cover all types of drugs used in any setting. Their function is to identify potential safety issues as soon as possible, and to continuously monitor and evaluate potential safety issues in relation to reported ADRs.[14]

The increase of ICSRs in SRSs has allowed the application of data mining and statistical techniques for signals detection. However, the success of signal detection in SRSs is hampered by the concept of voluntary reporting, including factors that may influence the reporting rate and quality of data, lack of an accurate quantification of the frequency of events or the identification of possible risk factors for their occurrence, and missing denominators. To overcome some of these limitations, the US FDA[15] and EMA adopted other sources of databases including EMRs and AHD. [16]

5.2 Healthcare System Databases

Health Care System databases may include the following:

- EHR: Refer to as "An electronic record of health-related information on an individual that conforms to nationally recognized interoperability standards and that can be created, managed, and consulted by authorized clinicians and staff across more than one health care organization." [17] These are complete health records under the custodianship of a health care provider(s) that holds all relevant health information about a person over their lifetime. This is often described as a person-centric health record, which can used by many approved health care providers or health care organizations. [18] EHR include all personal health information belonging to an individual; entered and accessed electronically by healthcare providers over the person's lifetime; and extends beyond acute inpatient situations including all ambulatory care settings at which the patient receives care."[19]
- EMR: Refers to as "an electronic record of health-related information on an individual that can be created, gathered, managed, and consulted by authorized clinicians and staff

[14] http://www.who-umc.org/graphics/27418.pdf
[15] http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM464043.pdf
[16] http://onlinelibrary.wiley.com/doi/10.1038/clpt.2012.54/pdf
[17] https://www.nachc.com/client/Key%20HIT%20Terms%20Definitions%20Final_April_2008.pdf
[18] https://www.infoway-inforoute.ca/en/what-we-do/blog/digital-health-records/6852-emr-ehr-and-phr-why-all-the-confusion
[19] http://www.wpro.who.int/publications/docs/EHRmanual.pdf

within one health care organization." [17] These are partial health records under the custodianship of a health care provider(s) that holds a portion of the relevant health information about a person over their lifetime. This is often described as a provider-centric or health organization-centric health record of a person.[18] EMR system include patient information including socio-demographic characteristics, medical and drug history, diagnoses information including laboratory results, treatments and outcomes.[20]

- AHD: This is generated through the routine administration of health care programs.[21] These databases include characteristics of individual members, basic demographic data, therapeutic procedures, treatments and outcomes, diagnoses of hospital admissions, reimbursed prescriptions of drugs, etc.[19]
- Registries: Contain standardized information about a group of patients who share the same condition or experience.

Although these databases are designed primarily for routine clinical care they have been frequently utilized for observational studies. Their representativeness of routine clinical care makes it feasible to study "real world" safety, effectiveness and prescription patterns. In addition, these databases have been used in pharmacovigilance predominantly to confirm signals from SRSs on an *ad hoc* basis.[22]

These databases are also an important source of data for detection of ADRs. However, extracting signals from these databases requires specific methods to analyse the unique issues that arise from working with non–ICSR data.[22] The figure below illustrates a possible approach on the utilization of EHRs and other similar systems to complement the signal detection from SRSs.[23] Natural language processing is used to extract potential drug-outcome interactions from unstructured component of EHRs (i.e., narratives), which are then linked to each structured EHR.[24] The MedLEE product is one example of currently available software that is capable of performing natural language processing on the unstructured text within EHRs.

[20] http://cdn.intechopen.com/pdfs/38579/InTech-Data_mining_techniques_in_pharmacovigilance_analysis_of_the_publicly_accessible_fda_adverse_event_reporting_system_aers_.pdf
[21] http://mchp-appserv.cpe.umanitoba.ca/viewDefinition.php?definitionID=102210
[22] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3685297/
[23] http://jamia.oxfordjournals.org/content/20/3/413
[24] Friedman, C., 2000. A broad-coverage natural language processing system. In *Proceedings of the AMIA Symposium* (p. 270). American Medical Informatics Association.

Unstructured EHR refers to EHR narratives (using the natural language processing system MedLEE). Structured EHR data refers to laboratory test results, which are linked to each narrative) AERS∩EHR=common to both systems.

5.3 Non-conventional Data Sources

Non-conventional data sources are primarily non-structured data that may be used to complement existing SRSs. Non-conventional data sources include the biomedical literature, social media (including Twitter, Facebook, Patient Forums) and web search terms. Further research is required to demonstrate the benefit of using these data in post-market signal detection

6 ICMRA Pharmacovigilance Sub-Group Members' Experiences

Section 6, 7 and 8 of this paper captured verbatim contributions from each member.

6.1 Australia

6.1.1 Big Data and Analytics Experiences

There is significant interest across the Common wealth Government of Australia in using big data and big data analytics, including data integration or linkage, to maximise the potential value of existing and new datasets. These approaches are seen to be a cost-effective and timely way of utilising data to improve government services, and can reduce the duplication of information collection from stakeholders.

The Australian Government Department of Health is involved in a cross-portfolio project to create a data warehouse that will facilitate big data analytics for health-related datasets held by the Department.

The use of Big Data in the field of pharmacovigilance in Australia is still in its infancy. The TGA is currently involved in a number of projects which seek to enhance our capability to utilise Big Data for the purposes of pharmacovigilance.

6.1.1.1 *Spontaneous Reporting*

Adverse event monitoring in Australia occurs primarily through a spontaneous reporting system administered by the TGA. Marketing authorisation holders are mandated to report serious adverse events to the TGA, and we also receive spontaneous reports from clinicians, consumers and State and Territory health authorities. Analysis of these reports is a significant source of signal detection within Australia.

As of April 2016, there are over 380,000 individual adverse event reports contained within the TGA Adverse Drug Reaction System (ADRS) database. As the volume of data encoding these reports is very small, the ADRS database would not meet any criteria for being Big Data. The potential for linkage of the data contained with the ADRS to other datasets is also very limited as adverse event reports are de-identified and often incomplete in terms of information that might be used for probabilistic linkage.

Disproportionality analysis of adverse event reports within the database is conducted periodically using the proportional reporting ratio (PRR). The TGA is currently involved in upgrading the platform on which the ADRS is built and the analytic tools available for use, to enable additional measures such as the Information Component and ROR to also be calculated, along with confidence intervals for the PRR. There will also be the potential to run ad-hoc statistical disproportionality queries.

The TGA sends extracts from the ADRS database to the WHO-UMC on a monthly basis. The TGA is exploring ways of improving the completeness of reports sent to the WHO.

6.1.1.2 *Health Care System*

The TGA is exploring ways in which pre-existing health care system datasets can be utilised for pharmacovigilance. The main goal in the initial stages of this work is to use these datasets to investigate signals detected through other means, rather than for signal detection. While the use of these datasets for signal detection remains a long term possibility, the capacity and infrastructure to undertake this work is being developed and confirmatory work on the ability of the datasets to identify signals first needs to be established as the datasets where not constructed for this purpose and may not cover all health care provided.

The two main health-related administrative datasets are those associated with the Medicare Benefits Schedule (MBS) and the Pharmaceutical Benefits Schedule (PBS). Both MBS and PBS datasets record financial transactions for either medical services in the case of the MBS, or medicine prescriptions in the case of the PBS. The MBS captures medical services that are funded by the federal budget, but does not capture services provided under state health budgets, such as public hospital services. The MBS data includes the information that a consultation with

a general practitioner or a specialist has occurred but does not indicate the subspecialty, the reason for consultation or the outcome. The MBS data also includes information on procedures, and pathology and imaging tests but generally does not include information on the condition being investigated or treated.

Likewise, the PBS dataset captures community-based dispensing of prescription medicines subsidised by the Australian Government, but does not capture information on medicines supplied by public hospitals or medicines not listed on the PBS. The indication for treatment is generally not recorded except in some circumstances where the medication requires a specific authorisation. While there are limitations to each dataset, both have the advantage of covering the vast majority of the Australian population.

As an indication of the scale of these datasets, the following table shows the number of services processed under the MBS and PBS for the past three financial years:

Services processed	MBS	PBS/RPBS
2014-2015	373.4 million	225.6 million
2013-2014	358.3 million	224.1 million
2012-2013	344.0 million	211.1 million

Source: Australian Government Department of Human Services, Annual Report 2014-2015 (https://www.humanservices.gov.au/sites/default/files/documents/8802-1510-ar2014-15.pdf)

Each dataset has unique properties relating to changes in governance and policy over time. These present challenges when designing research questions around them. The TGA recently conducted a data-linkage study combining MBS and PBS datasets to investigate a possible relationship between a medicine used to treat multiple sclerosis, and the development of malignant melanoma. The results of the project did not support an association but it did provide valuable insight into the challenges and opportunities associated with using these big datasets for pharmacovigilance purposes. One of the biggest challenges is in ensuring that the outcome being measured is a valid and reliable proxy for the outcome of interest, and in selecting medicine-adverse event pairs that will be adequately captured by the datasets. For example, the number of individuals who had a melanoma excision recorded in the MBS dataset after exposure to the medicine of interest was only 18. This was not unexpected given known epidemiological estimates for the prevalence of MS and the incidence of melanoma in Australia, however, it highlights the fact that Australia has a relatively small population which can reduce the statistical power available to researchers using these datasets.

Data concerning hospital admissions is collated by the Australian Institute of Health and Welfare, based on that provided by state and territory health authorities. A national minimum data set is provided by each jurisdiction, including episode-level records containing administrative, geographic and clinical data from admitted patient morbidity data collection

systems in Australian public and private hospitals. The TGA is exploring ways in which linking this data to PBS prescription data could be used for signal investigation.

The TGA is exploring other options for using other linked health datasets, including using the Sax Institute 45 and Up study (https://www.saxinstitute.org.au/our-work/45-up-study/) to investigate the feasibility of data linkage for signal investigation. The 45 and Up study population is approximately 10 per cent of New South Wales (NSW) residents aged over 45 years. While this is a small sample, consent has been given for the linkage of data from many different sources, including hospital admissions (including the NSW Admitted Patients Data Collection), general practice records, MBS, PBS, and various other health related registries. The particular advantage of this study is the access to diagnostic codes from hospital and general practice datasets, which should increase the validity of outcome measures. The 45 and Up study uses a purpose-built secure research environment which allows researchers secure, remote access to the anonymised linked data from the 45 and Up study.

The Australian Government Department of Health administers the My Health Record program which was rolled out nationally in July 2012. The My Health Record is a personally controlled electronic health record which contains health information relating to allergies, current conditions and treatments, medicine details, pathology reports or diagnostic imaging reports accessible by healthcare providers nationally. Currently, the system operates on an opt-in basis, however there is an opt-out pilot study being conducted for a small, geographically defined sample. The My Health Record is still in a relatively early stage of development with the development of responses to public consultation still ongoing. There are no plans or permissions currently for information stored within the My Health Record to be used for pharmacovigilance purposes, although it remains a potential tool for exploring hypotheses in future.

MedicineInsight is a primary care quality improvement program, run by the National Prescribing Service. It was established in 2011 and collects de-identified data directly from general practice electronic medical records. There are currently around 500 individual practices participating, with data collected on approximately 3.5 million patients. The objectives of the program are to improve the post marketing surveillance of medicine use in Australia, and the develop quality improvement activities in general practices. The data are currently being used to assess the uptake of new drugs, and evaluate compliance with clinical guidelines. The potential benefit of this data to pharmacovigilance activities is the ability to access clinical outcome information linked to prescribing information.

Registry information is currently in use for medical device vigilance. The Devices Vigilance team within the TGA is exploring options for linking the National Joint Replacement Registry to other datasets to enable better surveillance of clinical outcomes following implantation of orthopaedic prostheses. In the pharmacovigilance area, we do not routinely access registry information, although the possibility to link certain registries, including cancer and death registries, is available through the Sax Institute 45 and Up study.

The TGA is also exploring how data from the Australian Childhood Immunisation Register (ACIR) might be useful in investigating safety signals associated with vaccines. The ACIR is a national population-based register. It was established in 1996 and initially recorded details of vaccinations given to children under seven years of age who live in Australia. In January 2016 the register expanded to show the childhood vaccination history for adolescents up to the age of 20 years. In September 2016, the ACIR will become the Australian Immunisation Register (AIR). The AIR will enable the capture of all vaccines given to eligible Australians throughout their life.

TGA has not undertaken data analytics projects for non-conventional data sources, such as social media or patient forums for medicinal products, however, this has been undertaken as a small pilot for medical devices. While this is a future area that could be considered for development the major focus has been on the use of robust health datasets until more information about the validity of social media data is known. There is interest in learning from other agencies experiences in this area.

6.1.2 Opportunities

The use of Big Data approaches in the pharmacovigilance space provides us with new opportunities for turning routinely collected data into meaningful information. In the case of administrative datasets, finding uses for the data in addition to their primary purpose ensures that the population is able to realise the greatest possible value from it, in both health and financial terms. Other advantages include:

- The potential to complement traditional spontaneous reporting systems, by allowing an epidemiological approach to determining the incidence of adverse events in the population.
- Better ability to identify and investigate medicine-adverse event associations which occur over a longer time period, which may not be identified through spontaneous adverse event reports which depend on a close temporal relationship.
- Greater potential for the investigation of signals across different sub-populations and to control for confounders.

6.1.3 Challenges

There are many challenges which need to be overcome before Big Data approaches to pharmacovigilance become business-as-usual in Australia. As alluded to previously, the administrative datasets available to us have been collected for specific reasons which are unrelated to the purposes of pharmacovigilance. The main challenges include the following:

- As the biggest datasets are collected for financial reasons, the data are influenced by practitioners' behaviour around funding models.

- Data collected under the MBS record an item of service but little information regarding the clinical reason for the service, or the outcome of the service.
- The rules governing these datasets are subject to change over time. There is currently no electronic access to historical records of these datasets.
- Multijurisdictional custodianship of various datasets inhibits a streamlined approach to data access, even within the one governmental department.
- While case note information would be a valuable resource for pharmacovigilance, much of it is unstructured, and again, not recorded for the primary purposes of pharmacovigilance. The TGA currently does not have access to the infrastructure and resources necessary for the analysis and interpretation of such data if it were available.
- There is a need for greater internal capacity to be involved in Big Data projects into the future, such as expertise in study design, statistics, detailed knowledge of the datasets, and information technology.

6.2 Brazil

6.2.1 Big Data and Analytics Experiences

6.2.1.1 *Spontaneous Reporting*

NOTIVISA is the post-market surveillance electronic system that collects and assesses reports of suspected adverse reactions to health products (medications; natural health products; biologics [biotechnology products, vaccines, fractionated blood products, human blood and blood components, as well as human cells, tissues and organs]; radiopharmaceuticals; cosmetics; medical devices; and disinfectants and sanitizers with disinfectant claims) marketed in Brazil.

The Brazil Vigilance Electronic System has collected reports of suspected adverse reactions since 2007. Adverse reaction reports are submitted by health professionals and consumers on a voluntary basis either directly to Anvisa or via Market Authorization Holders (MAHs). In addition, MAHs are required to report serious adverse events (RDC 04/2009). Since 2007 until 2014, the Brazil Vigilance database (NOTIVISA) contained 225,078 reports for all health products, including 77,370 regarding medicines. These numbers are expected to increase with the implementation of mandatory electronic adverse reaction reports for Health Institutions (RDC 36/2013).

In Brazil, the approach of signal detection from spontaneous reporting has been the manual review of individual case safety reports followed by formulation of hypotheses, leading to further investigations which sometimes result in regulatory warnings and changes of the product monographs and in some instances withdrawals of marketing authorizations.

Development of data mining capacity in Brazil Vigilance database is being explored. Besides, ANVISA is studying the possibility to purchase a new system which includes data mining tools and open the expectation to link this new system to public systems available in the country.

Health Care System

The Sentinel Network Experience

Given the difficulty in receiving adverse reactions reports and technical complaints about the various health products, including medicines, there is damage in Anvisa activity since spontaneous reporting has not reached the volume and the degree of confidence desirable to base market regulation.

The Sentinel Network was created to respond to this need to get qualified information, while creating a favorable intra-hospital environment to the development of health surveillance in hospitals, which should result in significant quality gains for services and patients.

The Brazilian Network of Sentinel Hospitals, or simply Sentinel Network, has as one of its aims to serve as an active observatory's performance and safety of health products under sanitary surveillance, including medicines. Initially composed of teaching hospitals or complex, it was expanded and today there are other health institutions.

The selection criteria favored the choice of large and medium-sized hospitals, which perform wide range of procedures with the participation of varied and complex medical technologies and develop residency programs.

For network deployment and reinforcement were developed training workshops for health professionals: doctors, nurses, pharmacists, engineers, related hospital areas: Engineering and Maintenance, Hospital Pharmacy, Hemotherapy services for setting up the core of the Management of Health and Hospital Risk - GR.

Currently, the Sentinel Network is composed by 219 services, and corresponds to approximately 50% of all adverse events reports received for medicines. Regarding the Pharmacovigilance, besides receiving reports, the Network is a constant tool for information exchange, such as the reinforcement signal in searching for changes in the benefit-risk medication profile.

Other initiative is that the serious adverse events reports are mandatory for all hospitals since 2013 (RDC 36/2013).

6.2.1.2 *Non-Conventional Data Sources*

Brazil currently uses a tool to search for rumors regarding health products and some key words which can change depending on the aim of the search.

There are no other non-conventional data sources for signal detection yet (Twitter, Facebook, Patient forums, etc.).

6.2.2 Challenges

Challenges for combining different types of datasets include the following:

- Verifying appropriate datasets to enhance post-marketing pharmacovigilance.
- Bringing together different organizations domestically and internationally.
- Developing robust and consistent mechanisms to ensure patient privacy and appropriate use of data.
- Developing standards, principles, and best practices for visibility and accessibility of data.
- Building an infrastructure to support datasets and develop innovative methods to link together different types of datasets.
- Purchase statistic tools in order to enhance signal detection.
- Employing statistical algorithms to extract information from datasets and the appropriate statistical expertise to utilize the data.
- Mapping between different coding systems for health care data to extract ADRs.

6.3 Canada

6.3.1 Big Data and Analytics Experiences

6.3.1.1 *Spontaneous Reporting*

The Canada Vigilance Program is Health Canada's post-market surveillance program that collects and assesses reports of suspected adverse reactions to health products (prescription and non-prescription medications; natural health products; biologics [biotechnology products, vaccines, fractionated blood products, human blood and blood components, as well as human cells, tissues and organs]; radiopharmaceuticals; and disinfectants and sanitizers with disinfectant claims) marketed in Canada.[25]

The Canada Vigilance Program has collected reports of suspected adverse reactions since 1965. Adverse reaction reports are submitted by health professionals and consumers on a voluntary basis either directly to Health Canada or via Market Authorization Holders (MAHs). In addition, MAHs are required to report serious adverse events in Canada and unexpected serious adverse events internationally (C.01.0.17 of the Food and Drug Regulation). In 2015, the Canada Vigilance database contained 1,181,939 reports for all health products, including 500,000 domestic reports. These numbers are expected to increase with the implementation of mandatory electronic adverse reaction reports.

In Canada, the traditional approach of signal detection from spontaneous reporting has been the manual review of individual case safety reports followed by formulation of hypotheses, leading to further investigations which sometimes result in regulatory warnings and changes of the product monographs and in some instances withdrawals of marketing authorizations.

[25] http://www.hc-sc.gc.ca/dhp-mps/medeff/vigilance-eng.php

Development of data mining capacity in Canada Vigilance database is being explored. However, Health Canada has been able to utilize other resources to address a number of safety signals. These include communications with other foreign regulatory agencies, mainly those in Europe, United States, and Australia on a regular basis to determine whether similar adverse reactions have been reported in their jurisdictions. Health Canada also queries the World Health Organization (WHO) Uppsala Monitoring Center database, which collects adverse reactions data from over 130 countries participating in the WHO International Drug Monitoring Program, and has data mining capacity.[26] In addition, Health Canada maintains a database of drug safety information received from the European Medicines Agency's Pharmaceutical Risk Management Committee. This database contains outcomes of risk assessments conducted by the member states of the European Union.

6.3.1.2 *Health Care System*

Health Canada is continuously looking for new ways to enhance the quality and quantity of adverse reactions reports. To complement the Canada Vigilance Program, the following datasets may be considered: AHD, EMRs and registries established and maintained by the provinces and territories and other insurers.

Combination of these datasets could provide critical information on adverse reactions for enhanced post-market pharmacovigilance, including the implementation of data mining methods, which will speed up the process of signal detection and improve the signal prioritization process of potential safety signals.

One potential new source of Canadian AR data is the electronic medical record (EMR). Across Canada, the eHealth system is comprised of compatible data repositories and registries operated at provincial/territorial jurisdiction levels and sourced from point of care patient/provider encounters. Each jurisdiction's approach to electronic medical records is unique and the extent to which patient health information captured in EMRs can be leveraged for pharmacovigilance depends on: 1) its intended use, 2) the quality and quantity of ADR data captured, and 3) whether it is entered in a way that supports analysis or extraction from the EMR.

There are several terminologies used across Canada to encode clinical information in patient records in the health care system. None are completely aligned with MedDRA, the terminology used by the Canada Vigilance Program for coding medical terms (reaction term, indication and patient history) in adverse reaction reports.

Some Canadian EMR users are moving toward adoption of SNOMED CT for coding clinical concepts. Other medical terminologies used by clinicians for indexing data in Canadian EMRs include ICD-10-CA, and ICD-9 (diagnostic codes) and the Anatomical Therapeutic Chemical (ATC) classification system (drug names).

[26] http://www.umc-products.com/DynPage.aspx?id=3521

In addition, the "*Protecting Canadians from Unsafe Drugs Act*" (Vanessa's Law) was legislated in 2014. This new legislation was introduced with the objective to improve patient safety from health products and medical devices in Canada by increasing Health Canada's ability to:[27]

- strengthen safety oversight of therapeutic products throughout their life cycle; and
- improve reporting by certain health care institutions of serious adverse drug reactions and medical device incidents that involve therapeutic products.

6.3.1.3 *Non-Conventional Data Sources*

Health Canada currently does not use any non-conventional data sources for signal detection (Twitter, Facebook, Patient forums, etc.). Further research is required to demonstrate the benefit of using these data in post-market signal detection.

6.3.2 Opportunities

Combining information from different adverse reaction datasets will complement limitations of each dataset on its own and will provide more opportunities to conduct analysis of patient characteristics, health products and the adverse reactions involved.

Health Canada has explored several opportunities to enhance the quality and widen the scope of pharmacovigilance. These included the following:

- Health Canada partnered with the Canadian Institute for Health Research to establish and expand the Drug Safety and Effectiveness Network (DSEN). DSEN has the capacity to systematically conduct drug safety and effectiveness research.
- Health Canada uses IMS 360 to obtain drug utilization data from pharmacies and hospitals.
- Health Canada has access to various administrative datasets from the Canadian Institute for Health Information (CIHI) that could potentially be linked together.
- Other data sources include surveys and vital statistics data from Statistics Canada as well as surveillance databases from the Public Health Agency of Canada.
- Health Canada also has access to medication error information from the National System for Incident Reporting.
- A project is currently underway to aggregate data from several Poison Control Centres across Canada to determine the usefulness of these data in pharmacovigilance.
- Health Canada, working with The Canadian Primary Care Sentinel Surveillance Network (CPCSSN), has completed a proof-of-concept project to evaluate the completeness and quantity of adverse drug reaction data collected in Canadian primary care EMRs. This work generated evidence regarding the potential uses and deficiencies of national, primary care EMR data for pharmacovigilance activities.[28]

[27] http://laws-lois.justice.gc.ca/PDF/2014_24.pdf
[28] Canadian Primary Care Sentinel Surveillance Network Data Sourcing Pilot Project – Phase II

- Health Canada is also exploring other EMR systems e.g. IMS Evidence 360 that has demonstrated excellent mapping of codes to Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).
- Health Canada is in the planning stage of a collaborative study with Canada's Drug Safety and Effectiveness Network (DSEN) to evaluate the utility and cost-effectiveness of a commercially available social media monitoring platform for pharmacovigilance.

In addition, Health Canada is studying the feasibility of collaboration with the US FDA Sentinel Program.

6.3.3 Challenges

Challenges for combining different types of datasets include the following:

- Verifying appropriate datasets to enhance post-marketing pharmacovigilance.
- Ensuring capture of priority data (e.g., suspect product, reaction, outcomes) in discrete/table driven entry by EMR users to fully support drug safety monitoring.
- Bringing together different organizations domestically and internationally.
- Developing robust and consistent mechanisms to ensure patient privacy and appropriate use of data.
- Developing standards, principles, and best practices for visibility and accessibility of data.
- Building an infrastructure to support datasets and develop innovative methods to link together different types of datasets.
- Employing statistical algorithms to extract information from datasets and the appropriate statistical expertise to utilize the data.
- Mapping between different coding systems for health care data to extract ADRs.

6.4 Ireland

6.4.1 Big Data and Analytics Experiences

Similarly to the subgroup members, HPRA operates a national spontaneous reporting system for suspected ADRs with storage on our national database. Reports are also provided to the Eudravigilance database and to Vigibase. The HPRA also participates in signal detection activities as part of the EU network work-sharing initiative.

Irish academic institutions are engaged in valuable pharmacoepidemiological research. Such research may generate evidence which informs regulatory decision-for example with respect to measuring the effectiveness of risk minimisation measures. In addition, with the announcement of the individual health identifier, whereby each individual using health and social care services in Ireland is assigned a unique, non-transferable number, there is potential for new data linkages within the Irish healthcare system in the future.

Within the EU Network, there are various initiatives including the EMA Patient Registry Initiative, an initiative that was established to identify and evaluate existing registries and develop a methodological toolkit to facilitate the establishment of high-quality new registries if none provide adequate source of post-authorisation data for regulatory decision-making. PRAC is establishing an approach for measuring the impact of pharmacovigilance activities that will deliver data, information and knowledge on major specific product related actions, on key pharmacovigilance activities and decisions and on enabling factors. The strategy will determine the conceptual approach, principles, stakeholders, priorities, and support for the planning of collection of data, information, and knowledge.

We agree with other ICMRA members that data protection, legal constraints and confidentiality considerations as well as establishing relationships and infrastructure to be able to access and link data are amongst the challenges faced by regulators. Data quality and a lack of standardisation are also challenges relevant to interoperability of systems. Methodological rigour is of obvious importance to ensure the decision relevance of evidence generated. The latter is an area where regulators can influence particularly in the area of guidance development (see GVP Guideline on PASS, ENCePP guidance, ISPE guidance).

6.5 Italy

6.5.1 Big Data and Analytics Experiences

6.5.1.1 *Spontaneous Reporting*

In Italy, the spontaneous ADRs are collected through the National Network of Pharmacovigilance (Rete Nazionale di Farmacovigilanza - RNF), an extensive network throughout the national territory that includes 21 Regions, 20 Regional Centres of Pharmacovigilance, 162 Local Health Authorities, 105 Hospitals, the Poison Control centers across Italy, 51 Research Institutes and 992 Marketing Authorisation Holders and the Italian Medicines Agency (AIFA). The RNF is also operating in connection with the European network for pharmacovigilance EudraVigilance that collects in a single database all data provided at national level by the EU countries as well as extra EU countries. Finally, all reports are also provided to the WHO Uppsala Monitoring Centre database, Vigibase.

6.5.1.2 *Health Care System*

AIFA developed registries for monitoring the appropriateness of prescription and as a tool for risk-benefit and cost-efficacy. These structured data registries are designed primarily for HTA purposes however their representativeness of routine clinical care makes it feasible to study "real world" safety, effectiveness and prescription patterns. The AIFA Registers include more than one hundred medicinal drugs belonging to several therapeutic and specialized areas. The actors involved are: AIFA, the pharmaceutical companies, prescribers, hospital pharmacists, the regions, the National Health System (NHS).

6.5.1.3 *Non-Conventional Data Sources*

AIFA currently does not use any non-conventional data sources for signal detection (Twitter, Facebook, Patient forums, etc.).

6.5.2 Opportunities

The new internal regulation of AIFA features a new office, the "Database and analysis Office" born for the design and construction of the processing and interpretation of data systems (Big Data and Analitycs); the office will support the analysis and interpretation of data for the government of the Agency's activities and collaboration with national, regional and other international bodies in the determination of both the technical aspects of the content of information flows involving different cognitive and management activities of the Italian agency.

6.6 Japan

6.6.1 Big Data and Analytics Experiences

6.6.1.1 *Spontaneous Reporting*

Based on the PMD Act (revised Pharmaceutical Affairs Act)*, cases of Adverse Drug Reaction (ADR) and infections caused by drugs (both Japanese and foreign) are reported to PMDA by pharmaceutical companies. More than 98% of "serious" adverse events or reactions are reported electronically in ICH-E2B format and stored in PMDA's database. The number of the reports is indicated in Table 1.

Medical institutions (hospitals) also report cases of ADR and infections caused by drugs directly to PMDA based on PMD Act (and Preventive Vaccination Act in case of Vaccines). The format is defined by the Ministry of Health, Labour and Welfare (MHLW) and PMDA accept the report via e-mail, fax and shipping mail. These reports are stored in the PMDA's database together with the ones reported by pharmaceutical companies. The number of the report is indicated in Table 1.

PMDA utilizes the database for safety evaluation including signal detection (ROR and others).

Data set of such reports from companies (in Japan) as well as from medical institutions (called "Japanese Adverse Drug Event Report database (JADER)", in csv format) is published on PMDA's website https://www.pmda.go.jp/safety/info-services/drugs/adr-info/suspected-adr/0003.html (only in Japanese) and researchers can download and utilize it for academic research. Line listings (only in Japanese) of the reports are also viewable on PMDA's website.

ADR reports from patients have been accepted since March 2012, on a trial basis. These reports are sent to PMDA via PMDA'S web site. Number of the report is indicated in Table 2. PMDA plans to implement this system formally by FY 2018.

Table 1.

		2011FY	2012FY	2013FY	2014FY
Reports from Pharmaceutical Companies	adverse drug reactions (Japanese)	36,641	41,254	38,329	49,198
	cases of infections caused by drugs (Japanese)	100	159	98	78
	adverse drug reactions (foreign)	220,410	261,823	266,506	300,191
	cases of infections caused by drugs (foreign)	45	39	33	25
Reports from Medical Institutions (including Vaccine Adverse Reaction)		5,231	4,147	5,420	6,180

Table 2.

	2012.3.26～ 2012.3.31	2012FY	2013FY	2014FY
Reports from Patients	30	154	122	91

6.6.1.2 *Health Care System*

PMDA has initiated the Medical Information for Risk Assessment Initiative (MIHARI) Project since fiscal year (FY) 2009 with the aim of utilizing large-scale electronic health information databases as novel information sources for pharmacoepidemiological drug safety assessments in Japan. The meaning of "MIHARI" in Japanese is "to monitor" or "to watch over". In this project, more than 40 pilot studies have been conducted to characterize the databases for post-marketing drug safety assessment and to accumulate PMDA's experiences in conducting pharmacoepidemiological studies for which various designs such as cohort studies, nested case-control studies, sequence symmetry analyses (SSA), and self-controlled case series studies were employed. The major data sources used in the pilot studies are insurance claims databases, a database comprising electronic medical records (EMRs) from hospitals, and an inpatient care database. In addition, the regulatory guidance for pharmacoepidemiological studies using health information databases for drug safety assessments was published in 2014 to ensure the appropriate use of these databases by industries.

In FY 2011, PMDA in collaboration with Ministry of Health, Labor and Welfare (MHLW) has initiated to develop a new system called as "MID-NET"(Medical Information Database-Network) which is a distributed EMRs databases network among the 23 hospitals in Japan. In these databases, various linked-data including laboratory result values are stored and are automatically updated every week. PMDA is currently working on data/system validation and pilot studies to assure the integrity of the MID-NET system for full implementation in 2018.

More information (only in Japanese) is available on the PMDA website at the address below.

- MIHARI: http://www.pmda.go.jp/safety/surveillance-analysis/0011.html
- MID-NET: http://www.pmda.go.jp/safety/surveillance-analysis/0018.html

Example of studies conducted in MIHARI were shown below.

Takeuchi, Y et al, Atypical Antipsychotics and the Risk of Hyperlipidemia: A Sequence Symmetry Analysis. *Drug Saf* 38: 641-650 (2015).

Ishiguro, C et al. The MIHARI Project: Establishing a new framework for pharmacoepidemiological drug safety assessments by the Pharmaceuticals and Medical Devices Agency of Japan. *Pharmacoepidemiol Drug Saf* (in press)

6.6.1.3 *Non-Conventional Data Sources*

Currently, PMDA* does not utilize social media as data source for assessment.

6.6.2 Opportunities

It will be beneficial to share regulatory experiences about utilization of health record information database for understanding current situations and limitations on those utilization for regulatory purpose, and promoting use of pharmacoepidemiological approaches in regulatory setting. Such activities can identify tasks which should be discussed internationally among regulatory authorities and may also facilitate international cooperation and collaboration among regulatory authorities in this field. Especially, drug safety in post market stage is common interest among regulatory authorities. Thus, international guideline on utilization of medical health record for drug safety assessment will be necessary in the near future for promoting international harmonization and proper use of data by industries in pharmacovigilance.

6.6.3 Challenges

Situations and progress may vary among regulatory authorities, including not only differences on methods for data standardization and coding techniques but also differences in resources and in a medical infrastructure including healthcare insurance system.

* Tentative translation of the PMD Act (revised Pharmaceutical Affairs Act) is as follows: the Act on Securing Quality, Efficacy and Safety of Pharmaceuticals, Medical Devices, Regenerative and Cellular Therapy Products, Gene Therapy Products, and Cosmetics.

6.7 Netherlands

6.7.1 Big data in the Netherlands

In the Netherlands, the usage of electronic medical records is high, and still increasing. This means that potentially there is a wealth of 'big data' available. There are several independent organisations and research institutions throughout the country who have created, and maintain and further expand databases that are being used for scientific research (see table 1 below). Although there is no central data capture

point, and the MEB does not 'own' one of the database, the MEB is very much interested in those data and what these can add to evaluation of benefits and risk of medicines, and therefore there are close collaborations with the vast majority of these datasets.

ADR monitoring in the Netherlands is maintained by the Netherlands Pharmacovigilance Centre Lareb on behalf of the MEB. Marketing authorisation holders are mandated to report serious adverse events to Lareb. Moreover Lareb collects spontaneous reports from health care professionals and consumers, maintains the national database, carries out signal detection on national data and reports any signals to the MEB. All reports are also provided to the Eudravigilance database, operated by the European Medicines Agency (EMA) and the WHO Uppsala Monitoring Centre database, Vigibase.

The common challenges on security and privacy, partnerships, infrastructure and capacity, and standards mentioned in the executive summary also apply to the situation in the Netherlands. The Dutch healthcare system is organised as a sentinel system with the GP having access to the medical records of patients including information of i.e. hospital care.

6.8 New Zealand

6.8.1 Big Data and Analytics Experiences

6.8.1.1 *Spontaneous Reporting*

Medsafe is the New Zealand authority responsible for ensuring that medicines approved for use in New Zealand are of acceptable efficacy, quality and safety. The Centre for Adverse Reactions Monitoring (CARM) as part of the New Zealand Pharmacovigilance Centre (NZPhVC), is contracted by the Ministry of Health (Medsafe) to collect, collate and analyse suspected adverse reactions to therapeutic products (medicines, vaccines and complementary products) that are available in New Zealand. Medsafe receives information regularly from CARM about spontaneous adverse reactions and uses this information in their decision making process.

CARM has collected reports of suspected adverse reactions since 1965. Adverse reaction reports may be submitted by anyone, including pharmaceutical companies, healthcare professionals and consumers. From inception of the spontaneous reporting scheme until the end of April 2016 almost 96,000 (95,436) reports have been submitted. This includes natural health products/complementary products.

Pharmacovigilance guidelines, developed by Medsafe, are part of current guidelines on the regulation of therapeutic produces in New Zealand. These guidelines includes recommendations to pharmaceutical companies that are not currently underpinned by medicines legislation. These recommendations aim to provide guidance on best practice in terms of pharmacovigilance (www.medsafe.govt.nz/regulatory/Guideline/GRTPNZ/part-8-pharmacovigilance.pdf).

Disproportionality analysis of adverse reaction reports within the CARM database is conducted on a monthly basis by Medsafe using the proportional reporting ratio (PRR).

CARM sends extracts from the database to the WHO-UMC on a monthly basis. CARM uses WHO-ART which is mapped to MedDRA. ICD-10 is used for administrative health data (eg, diagnoses) and ATC groupings are used for medicines.

6.8.1.2 *Health Care System*

The New Zealand health care system is predominantly public funded. Every eligible person is assigned a unique National Health Index (NHI) number, which is stored in a database along with that person's demographic details. This helps with the planning, coordination and provision of health and disability support services across New Zealand.

The NHI number ties together all patient information and patient systems, and is used to report patient events to national data collections such as those held by CARM and the Ministry of Health. The Ministry of Health records events based on NHI number and holds anonymised statistical databases. These databases are used to meet reporting requirements, develop policy, facilitate research, monitor performance of services and support the planning of services. It is possible to extract some information from these databases.

The national Medical Warning System (MWS) is designed to warn health and disability support services of any known risk factors that may be important when making clinical decisions about individual patient care. The MWS is associated with NHI. CARM has privileged access to the MWS to manually enter alerts as Warnings (Precautions) or Dangers (Contraindications).

PHARMAC, the government pharmaceutical management agency, decides which pharmaceuticals to publicly fund in New Zealand. The majority of medicines that are dispensed through community pharmacies are funded through this system and the patient only pays a small co-payment. This data is collected and dispensing numbers from community pharmacies can be obtained (usage data for a defined part of the population, ie those getting medicines via prescription).

Some parts of the electronic health record are integrated, whilst there is work continuing to incorporate areas that are separate.

6.8.1.3 *Non-Conventional Data Sources*

Medsafe does not currently use any non-conventional data sources for signal detection.

6.8.2 Opportunities

The opportunities identified by other countries are also relevant to New Zealand. There is the potential to combine various datasets to obtain more in-depth information to enhance the quality of information available. Better identification of adverse event data would be beneficial and this could complement existing methods of collecting spontaneous adverse reaction information.

Medsafe also works with the Medication Error Reporting Programme (MERP), which is also part of the NZPhVC. The MERP and Medsafe also work in conjunction with the Health Quality and Safety Commission to reduce patient harm, such as through the safe use of medicines.

6.8.3 Challenges

Similarly, some of the challenges identified by other countries are also seen as a challenge in New Zealand. The main challenge is integration of different information systems (methods are needed to link together different types of datasets). This includes standardisation of data and coding techniques.

Another challenge is the custodianship of various datasets. There are different reasons for the collection of various information, which is unrelated to the purposes of pharmacovigilance (eg, medicines funding).

6.9 Singapore

6.9.1 Big Data and Analytics Experiences

6.9.1.1 *Spontaneous Reporting*

Health Sciences Authority (HSA), Singapore, is the national PV Centre which closely monitors the benefit-risk profile of all marketed health products in Singapore. The Vigilance and Compliance Branch, HSA, reviews all spontaneous adverse event (AE) reports from medical doctors, dentists, pharmacists, nurses and pharmaceutical companies. The total population in Singapore was 5.6 million in 2016. From 1993 to September 2016, the total number of AEs received by HSA is 234,500. Singapore is ranked first in terms of the number of valid individual case safety reports per million inhabitants submitted to the World Health Organization's global Database since 2011. The annual AE reporting rate for Singapore is more than 3,500 reports per million inhabitants.

The majority of the AE reports are from healthcare professionals who report suspected AEs on a voluntary basis and sent through the Critical Medication Information Store (CMIS). The CMIS is a national electronic platform in all public healthcare institutions in Singapore, which allows healthcare professionals to record, access adverse drug reactions in the patients' medical records online, and submit these reports directly to HSA. It is also legally mandatory for pharmaceutical companies to report AEs to HSA although these reports account for a smaller fraction of the total no of reports. The AE reports are codified according to the WHO Adverse Reaction Terminology (WHO-ART). For the purpose of data analyses, WHO preferred terms are used for comparisons. HSA is adopting a new database to capture the AE reports and intends to use Medical Dictionary for Regulatory Activities (MedDRA) terms for coding the AE reports.

Quantitative signal detection methods, interpreted alongside detailed clinical review of reports, can be a valuable tool for national pharmacovigilance centres with small spontaneous AE reports database to triage, detect and assess safety signals. HSA is exploring the use of ROR, BCPNN, GPS and SPRT to detect potential signals of disproportionate reporting. There are some AE reports which require clinical review irrespective of the results of the quantitative data mining methods, these include: reports with fatal outcome and serious AE reports associated with traditional medicinal products.

6.9.1.2 *Health Care System*

National Electronic Health Record

A large volume of clinical data is captured in hospitals and outpatient clinics as part of clinical care. In Singapore, all public hospitals and government-supported outpatient clinics have electronic medical record systems for clinical management of patients, which include pharmacy records, laboratory tests, and discharge summaries.

To improve healthcare quality for all residents, increase patient safety, lower healthcare costs and develop more effective health policies, the Singapore's Ministry of Health (MOH) created the National Electronic Health Record (NEHR) vision – "One Singaporean, One Health Record" – that allows patient health records to be shared across the nation's healthcare ecosystem. The linking of patient data across the different databases is facilitated by the National Registration Identity Card (NRIC) number, whereby each Singaporean/Permanent Resident has a unique identity number linked to each individual. In 2009, the MOH Holdings (MOHH) initiated a project to develop the NEHR system to extract and consolidate the patients' record into one record. Authorised healthcare professionals i.e. doctors and healthcare staff who are involved in the patients' treatment, have access to information on a patient's condition and medication list. Methodology for anonymising the NRIC and other personal identifiers in the NEHR is under development for secondary data analysis while preserving patient privacy. The adoption of NEHR in the private sector is low but expected to increase over time.

Vaccine intensive monitoring programme

In 2009, the Vigilance and Compliance Branch of HSA partnered with a local children's hospital to conduct active surveillance for adverse events following immunization (AEFI) after influenza vaccination, as part of vaccine safety monitoring following pandemic influenza A (H1N1) public vaccination campaign. This programme was extended to include active surveillance for all vaccines given in childhood. A surveillance coordinator would screen all paediatric admissions for possible relationships to vaccination, excluding elective admissions and performed causality assessment for each case. Hospital-based active surveillance can enhance signal detection and

follow-up investigations of AEFI. A paper of their work was published and it reported a 5-fold increase (95% CI 1.2–33.1) in BCG-associated regional lymphadenitis.[29]

6.9.1.3 *Non-Conventional Data Sources*

Increasing numbers of patients have turned to social media to share their experiences with drugs, medical devices, and vaccines. Potential drug safety signals generated from social media data may need to be validated using more traditional pharmacoepidemiological methods. However at this point, HSA is not monitoring social media for pharmacovigilance.

6.9.2 Opportunities

HSA has embarked on a pilot project in 2014 to tap on the EMR of one public restructured hospital with the objective of establishing a national active surveillance network that leverages on EMR capabilities to identify patterns and early indications of ADRs in Singapore. The types of information include discharge summaries, laboratory records, inpatient and outpatient medications records. HSA is collaborating with data analytics, statistical and clinical experts to explore text and data analytics methods to pick out potential drug-AE associations.

6.9.3 Challenges

There are many diverse data environments and various formats in which patient information are being collected. While the NEHR system was rolled out in 2011, the system is mainly used by the public hospitals and government-supported outpatient clinics. For the private healthcare sector, the government is currently looking into ways to include them in using the NEHR system in the near future.

Due to the complexity of data stored in the various modules, the information extracted into NEHR may also not be in the consistent codified system. Additional cleaning of data (e.g. converting free text data to structured data manually) is both tedious and time-consuming.

A substantial amount of funding is needed to drive this initiative, especially if the goal is towards implementing a common data model to standardise data across the various institutions. Big data analytics requires a multidisciplinary team of experts such as data scientists, biostatisticians, clinicians, pharmacists, epidemiologists, research assistants, project managers and lawyers in order to achieve meaningful results and outcomes. Time needed from forging collaboration with various partners to implementation is substantial. Access to patient data in NEHR is also very restricted and require clearance from the various data owners.

[29] *Thoon KC, Soh SB, Liew WK, Gunachandran A, Tan NW, Chong CY, Yung CF. Active surveillance of adverse events following childhood immunization in Singapore. Vaccine. 2014 Sep 3;32(39):5000-5.*

6.10 South Africa

6.10.1 Big Data and Analytics Experiences

6.10.1.1 *Spontaneous Reporting*

The South African Pharmacovigilance deals with post-marketing surveillance program that collects and assess reports of suspected adverse reactions to medicines, which include prescription, over the counter medication, and vaccines. The South African Pharmacovigilance electronic system has 31585 individual case safety reports collected since 1992 to date. This number is expected to increase as the scope of the regulator is broadened to include regulation of medical devices, in vitro diagnostics, blood and its components as the regulator migrates into a new regulatory structure (the South African Health Products Regulatory Authority, SAHPRA). Previously adverse reactions were captured on the ADRI database. South Africa has recently (October 2015) acquired the WHO Vigiflow database, which feeds into the Uppsala Monitoring Centre database, Vigibase.

Adverse reaction reports are submitted by healthcare professionals and consumers on a voluntary basis either directly to National Adverse Drug Events Monitoring Centre (NADEMC), a pharmacovigilance satellite office of the Medicines Control Council or via Market Authorisation Holders (MAHs). MAHs are required by law to report all serious and non-serious, suspected adverse drug reactions, whether expected or unexpected, occurring in South Africa. MAHs should also report published accounts of suspected adverse drug reactions related to the active substance(s) of their medicine. All suspected adverse drug reactions from post-registration studies taking place in South Africa must be reported.

In South Africa, signal detection from spontaneous reports is conducted through the manual review of individual case reports from the line-listing followed by hypothesis formulation. Further investigations are conducted which may results in amendments in the package inserts (SmPC) and issuing of regulatory warnings such as Dear HealthCare Professional Letters (DHCPL) and in some instances withdrawals of product registration.

6.10.1.2 *Health Care System*

In South Africa, a pregnancy exposure registry/birth defect surveillance system has been developed and will be expanded to include other provinces. This registry aims to assess the safety of medicines used in pregnant women in terms of both maternal and neonatal outcomes at birth. The registry is currently in its third year of data collection.

Cohort-based pharmacovigilance studies have been conducted using large-linked databases in both the private and public sectors, particularly in the field of HIV. An approach aimed at collecting drug use and outcome data from the provincial electronic registration system and electronic pharmacy dispensing system is currently being explored in the Western Cape Province. An electronic ADR reporting application is being developed for inclusion into a

mobile phone application that provides health personnel across the country with primary health care treatment guidelines and HIV/TB management guidelines.

Targeted spontaneous reporting systems have been developed for HIV and TB medicines as part of the roll-out of antiretroviral treatments. This has increased reporting dramatically in South Africa. In order to reduce duplication of efforts and to harmonise the PV approaches employed by public health programmes, the essential medicines programme and the medicine regulatory authority, a collaborative working group has been established and meets on a regular basis.

The National Adverse Drug event Monitoring Centre works closely with the Medicines Information Centre in encouraging reporting of ADRs among its clients.

A decentralized system of ADR reporting, that involves more immediate and direct feedback to reporters has been launched in various provinces for HIV/TB medicines.

6.10.1.3 *Non-Conventional Data Sources*

South Africa does not utilise social media as data source for pharmacovigilance.

6.10.2 Opportunities

- this will complement the spontaneous reporting system and thus allowing an epidemiological approach in determining the incidences of adverse reactions
- potential for investigation of signals across the globe.
- enhanced surveillance system
- Building analytical capacity within regulatory authorities
- Detection of rare and long term ADRs

6.10.3 Challenges

- differences in data standardisation and coding techniques
- robust mechanisms to ensure patient privacy

6.11 Sweden

6.11.1 Big Data and Analytics Experiences

6.11.1.1 *Spontaneous Reporting*

In Sweden, the spontaneous ADRs are collected through the reporting from national Health Care providers e.g. physicians, nurses, dentists and pharmacists. There is also a consumer reporting system set up contributing to spontaneous ADRs.

The MPA is also providing ADR-data in relation to spontaneous ADRs within the European network for pharmacovigilance i.e. EudraVigilance. Finally, all reports are also provided to the WHO database Vigibase.

6.11.1.2 *Health Care System*

In Sweden there are, besides different medical record systems, a number of quality registries for the monitoring of the appropriateness of prescription and as a tool for the analyses of risk-benefit and cost-efficacy of drugs. The structured data registries are designed primarily for the Health Care system purposes. It is however also possible to use some of the available data to study pharmacovigilance that is "real world" safety with the actual drugs. The registers include data related to drugs belonging to different therapeutic and specialized areas. Currently there is ADR-reporting from some of the registries to the MPA.

6.11.1.3 *Non-Conventional Data Sources*

The MPA currently does not use any non-conventional data sources for reporting or signal detection.

6.11.2 Opportunities

The MPA has taken initiatives to gain access to safety data related to drugs available at other national regulatory bodies in order to increase the data for signal detection.

A project is also currently ongoing aiming at setting up electronic reporting directly to the MPA from the Health Care medical record systems. This will probably increase the possibility to have access to more safety data.

6.11.3 Challenges

The technical aspects with many different medical record systems and different technical set-ups for the quality registries are challenges to meet.

6.12 Switzerland

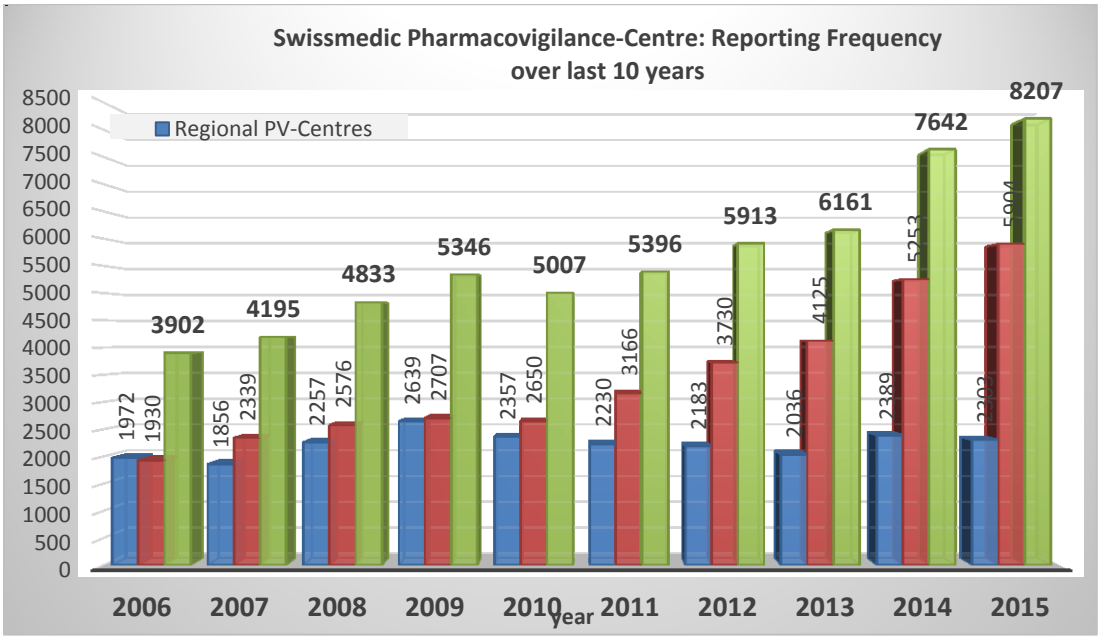
6.12.1 Big Data and Analytics Experiences

6.12.1.1 *Spontaneous Reporting*

Within the framework of the Swiss pharmacovigilance network, many of the reports on adverse drug reactions are received and assessed by six regional pharmacovigilance centers (RPVC) on behalf of Swissmedic and recorded in the national database in an E2B compliant format. Professionals who submit the reports receive appropriate feedback from the regional centers. Healthcare professionals can use the online reporting portal ElViS (Electronic Vigilance System) that was launched in October 2014 to report ADRs online to one of the regional pharmacovigilance centers. In 2015 Swissmedic received 115 reports from healthcare professionals via the portal. However, the majority of reports on suspected adverse reactions

from Switzerland are submitted to Swissmedic by pharmaceutical companies. Roughly 70% of the reports from companies were submitted during 2015 to Swissmedic electronically, using the pharmacovigilance gateway. Pharmaceutical companies without gateway access to the Swissmedic database can also send their reports to Swissmedic electronically via ElViS.

The number of ADR reports received by Swissmedic increased continuously during the past 10 years.



In 2015 Swissmedic received 8,247 reports of suspected adverse drug reactions: 2,307 of them were sent by the six regional pharmacovigilance centres (RPVC) and 5,940 by the industry. As in previous years, there was a sharp rise in the number of reports received (7.1%), due mainly to an increase in the volume reported by companies.

As of August 2016, the pharmacovigilance database of Swissmedic contains over 83,000 ADR reports originating in Switzerland. Since the volume of data encoding these reports is small, the Swiss ADR database would currently not meet any criteria for Big Data.

6.12.1.2 *Health Care System*

According to the Swiss Academies of Arts and Sciences the potential of Big Data is not fully exploited in the Swiss health care system. One of the reasons is that not enough relevant data is available.

This statement is also in accordance with the 2nd OECD report about the Swiss health care system as it recommends to invest more in acquiring health care information. The report goes on by specifically claiming that more effort is needed to collect data about treatment results and morbidity in relation to health care so that political decision makers know which are the biggest health risks and who in the population is most exposed to them.

As a consequence thereof, a federal law on national electronic health records was passed by the parliament in 2015. This law builds the basis for enhancing the Swiss health care system. It foresees the establishment of electronic patient records for each person living in Switzerland. These records can be made available to medical experts by the patient. (In case of an emergency where a patient is not responsive, medical experts can still access these electronic health records as long as the patient has not previously denied this option.)

With the accessibility of the national electronic health records always and everywhere the quality, efficacy and safety of the treatment will improve. The federal law on electronic patient records will become effective in 2017. Hospitals are given 3 years thereafter to start participating in the new system; care homes for the elderly are given 5 years. Once the national electronic health records are instituted, Big Data Analytics may be applied to improve the Swiss health care system even further.

More information (in German, French and Italian) is available on the link below:

http://www.e-health-suisse.ch/umsetzung/00135/00218/00256/index.html?lang=de

White paper of the Swiss Academies of Arts and Sciences (executive summery is available in German, French and English)

www.satw.ch/publikationen/communication1002.pdf

Assessment and recommendations of the OECD report 2011(in English)

http://www.bag.admin.ch/themen/internationales/11287/11326/13099/index.html?lang=en&download=NHzLpZeg7t,lnp6I0NTU042l2Z6ln1ad1IZn4Z2qZpnO2Yuq2Z6gpJCKdoB,e2ym162epYbg2c_JjKbNoKSn6A--

OECD Reviews of Health Systems - SWITZERLAND 2011 (in English)

http://www.ub.unibas.ch/digi/a125/sachdok/2012/BAU_1_5753611.pdf

e-Health
http://www.e-health-suisse.ch/umsetzung/00135/00218/00256/index.html?lang=de

6.12.1.3 *Non-Conventional Data Sources*

Currently, Swissmedic does not utilize social media as data source for pharmacovigilance assessment.

6.12.2 Opportunities

Generally, the Big Data approach provides a series of important opportunities in extracting valuable and significant information from large amounts of recorded datasets. In the case of pharmacovigilance, most important opportunities include:

- The potential to complete traditional reporting systems by adding the epidemiological approach in signal detection and risk evaluation.
- Setup of studies (including long-term observational studies) in order to confirm/reject potential drug-adverse reaction relationships.
- Increased versatility of analyses as well as enhanced statistical power and accuracy in various drug safety evaluations.

6.12.3 Challenges

In Switzerland, the major challenges to effectively using large amounts of Big Data can be summarised as follows:

- Legal: Massive restrictions in data-sharing/exchange due to privacy, data protection and security laws
- Technical infrastructure: A (global) platform capable of capturing large amounts of data from multiple sources and storing in a common data model does not currently exist.
- Methodological: Various coding and standardisation in currently existing datasets would not allow immediate and systematic use of available 'raw' Big Data.
- Resources: Much higher amounts of resources – including human, financial, etc. – would need allocation in Big Data projects.

6.13 UK

6.13.1 Big Data and Analytics Experiences

The MHRA is involved in a number of projects and initiatives dealing with "Big data" all with a view to harness data with a view to understanding what information exists on medicines so that we can be better informed on the benefits and risks. The UK government has a particular interest in innovation and the MHRA's digital and data strategies very much feed into the innovation agenda.

6.13.1.1 *Spontaneous Reporting*

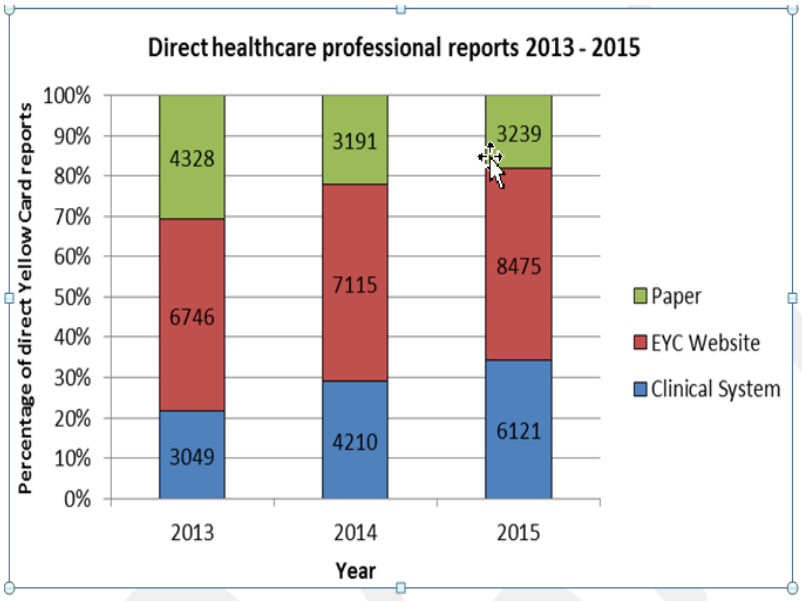
Adverse Drug Reaction (ADR) monitoring in the UK occurs primarily through a spontaneous reporting system administered by the MHRA and the Commission for Human Medicines (CHM), this is called the Yellow Card Scheme. Marketing authorisation holders are mandated to report serious adverse events to the MHRA, and we also receive spontaneous reports from healthcare professionals and patients. Analysis of these reports is the primary source of signal detection within the UK. All reports are also provided to the Eudravigilance database, operated by the European Medicines Agency (EMA) and the WHO Uppsala Monitoring Centre database, Vigibase.

As of April 2016, there are over 800,000 individual UK ADR reports contained within the MHRA's pharmacovigilance database. The annual volume of reports is increasing due to strategic initiatives to improve ADR reporting in the UK. The MHRA also collect worldwide data from manufacturers for signal detection purposes. We also have over 1.8M non-UK reports in our database.

As a result of a project that developed a standard (reference) for an E2B compliant ADR reporting form to be included in electronic GP systems and other clinical systems the annual reporting figures have increased greatly in the UK (see figure 1). The major difficulty in this

project has been the difference in terminologies used in the national healthcare system and the medicines regulatory system; to address this the MHRA completed a mapping of MedDRA terms to SNOMED-CT terms so that the data we receive through these channels can go directly into our database without the need for manual re-coding. This project is discussed in more detail in the paper on improving ADR reporting.

Figure1.



6.13.1.2 ***Health Care System***

As well as the electronic Yellow Card mentioned above the MHRA also operate and manage the Clinical Practice Research Datalink (CPRD and formerly known as GPRD). This is a live dataset of around 5 million active patients' records (8% of the UK population) and is increasing each year. The dataset is primarily used for public health research and pharmacovigilance activities. The PV department are working increasingly closer with CPRD to evaluate signals from spontaneous data, investigate the effectiveness of risk minimisation measures and understand the impact of regulatory interventions.

CPRD has provided a successful model in establishing the safety of vaccines by looking at "observed over expected" ADRs on HPV and Pertussis.

6.13.1.3 *Non-Conventional Data Sources*

6.13.2 Opportunities

WEB-RADR

Through the Innovative Medicines Initiative (IMI) MHRA is leading a large consortium of regulatory, industry, academic and technical partners in a 3 year project called WEB RADR (Recognising Adverse Drug Reactions). The project has four core objectives, (i) delivery of a mobile App platform for the reporting of suspected ADRs, (ii) harnessing social media data for information to benefit pharmacovigilance (iii) researching the benefits or otherwise of these data sources, and (iv) recommending policy around the application of these technologies and data in PV.

Through our consortium the project has already delivered a mobile App in the UK and the Netherlands and has gathered Facebook and Twitter data on 119 products for analysis. We are just over half way through the project and have already identified some potential applications of social media data as well as some early concerns. The research that will be carried out over the next 14 months will be key to regulatory understanding of what, if any, the role these "big data" can play in pharmacovigilance.

In the policy setting area the issues around data protection and ethical use of data will be addressed so that users of social media data understand the implications and constraints.

See http://www.WEB-RADR.eu for more information.

7 Security and Privacy

7.1 Australia

Within Australia, personal information collected by the Commonwealth Government, as well as certain private sector and not-for-profit organisations, is subject to the Commonwealth Government Privacy Act 1988, and the supporting Australian Privacy Principles which are found in Schedule 1 of the *Privacy Amendment (Enhancing Privacy Protection) Act* 2012. Health-related information is classified as sensitive information which is subject to more stringent controls.

The process of linking multiple datasets includes the possibility of creating potentially identifiable unit records. Therefore, agencies undertaking high risk data integration projects involving Commonwealth data for statistical and research purposes must be accredited. The accreditation process is conducted by the Cross Portfolio Data Integration Oversight Board against the interim accreditation scheme. There are currently three agencies within the Commonwealth that have this accreditation – the Australian Bureau of Statistics, the Australian

Institute of Health and Welfare, and the Australian Institute of Family Studies. The accreditation process for integrating authorities examines an agency's ability to ensure secure data management, technical capability and how the disclosure of identifiable information will be prevented.

7.2 Brazil

We based our decisions in two normative: Portaria 1660/2009 and the Information Access Law 12.527/2011. Those normative ensure that some data is confidential such as the patient and reporters identities and personal data.

7.3 Canada

Canada has a number of datasets that could be used for metadata analyses for pharmacovigilance, but the challenges of pooling different datasets are hindered by privacy and other laws across Canada. The privacy and security considerations of accessing health and health-related data in Canada were recently reviewed in depth by an expert panel of the Council of Canadian Academies.[30] Some relevant challenges are summarized below:

- Individual-level data held in different databases are not easily compared because they are not collected in a standardized manner;
- Pooled data analysis, which involves physical transfer of individual-level data to a central server, is often hindered by interpretation and implementation of privacy and other laws;
- The ability to access and link data within reasonable timeframes is uneven across Canada, or even lacking. Current rules and procedures to authorize research and allow data access overlap and are often time-consuming, processes and requirements for access are sometimes unclear, and access decisions are not always consistent; and
- Canada's governance of research ethics is fragmented, with significant differences across the provinces/territories. As well, laws on sharing data across provinces/territories and between countries differ or are lacking.

As 'Big Data and Analytics' (BDA) progresses within the Canadian health care system, challenges, risks and concerns arise about privacy, ethics and the legal implications related to BDA. These risks and challenges represent significant barriers and potential drawbacks, which may impact the success, pace and rate of adoption of BDA in Canada and the potential benefits it can bring.

In the context of the Canada Vigilance Program, personal information is collected pursuant to section 4 of the *Department of Health Act*, for the purpose of monitoring licensed products, detecting potential emerging safety issues and trends, mitigating the risks and improving the safe use and efficacy of the health products. Information related to the identity of the patient and/or

[30] http://www.scienceadvice.ca/uploads/eng/assessments%20and%20publications%20and%20news%20releases/Health-data/HealthDataFullReportEn.pdf

reporter will be protected as personal information under the *Privacy Act*, and in the case of an access to information request, under the Access to Information Act.

There are currently no Memoranda of Understanding (MOU), information sharing or technical agreements setting out the rules under which recurring and ad hoc disclosures of EMR information is permitted between jurisdictions or from jurisdictions to Health Canada for health product safety surveillance and monitoring activities.

In addition to the general legal context summarized above, there are additional internal policies that apply to Health Canada's collection, storage and handling of information that could include personally identifiable information or confidential information provided to Health Canada by third parties. Internal certification processes have been developed to ensure that all existing legal obligations are met, as well as compliance with Canada's Policy on Government Security. These processes include the performance of a privacy impact assessment, a threat risk assessment of the system architecture, and periodic auditing to ensure ongoing compliance. [31]

7.4 Ireland

7.5 Italy

In Italy, any personal information collected by the Public Authorities is subject to the Italian Privacy Regulation (D. Lgs. 196/2003) which introduces a Data Minimisation Principle. According to this principle, "Information systems and software shall be configured by minimising the use of personal data and identification data, by using either anonymous data or suitable arrangements to allow identifying data subjects only in cases of necessity, respectively".

Since health-related information are classified as sensitive information, two internal legal persons (data controller and data processor respectively) should ensure confidentiality and security of the processing of personal data and the relevant means, including security matters. Involved stakeholders in the Pharmacovigilance Data Management must be identified and accredited. The accreditation process is conducted by an Internal Board of the Agency according to an accreditation process with a different accessibility profile setting taking into account the nature of the stakeholder.

7.6 Japan

Regulation for personal data in Japan varies by the owner of information and consists of the following four legislations in general.

- *Act on the Protection of Personal Information* (No 57, Enacted in 2003);

[31] http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=16578

- *Act on the Protection of Personal Information Held by Administrative Organs* (No58, Enacted in 2003);
- *Act on the Protection of Personal Information Held by Incorporated Administrative Agency* (No59, Enacted in 2003); and
- *Act on the Protection of Personal Information Held by Local Assemblies*.

As seen above, different regulations are applied depending on the owner. However, it is generally possible to be applied to legalized business and consent personal matters.

In the meantime, in order to generate new industry and service together with an aim to improve public safety and reliability and protect personal information, revision is now considered in the law to be enacted in September 2017. The points of improvement can be noted as the following:

- Sensitive personal information requiring special attention can no longer be provided by opt-out.
- Usage of fabricated masked information with its intention not to specify particular individual shall be facilitated.
- Personal Information Protection Committee shall be established.

Necessary measures towards enacting the revised law are considered under the following aspects:

- Clarify the area secured for personal information which requires special attention
- Discuss actual methods to process fabricated masked information

Regulation to include items for fabricated masked information in the 1) *Act on the Protection of Personal Information Held by Administrative Organs*, and in the 2) *Act on the Protection of Personal Information Held by Incorporated Administrative Agency* is now put before the Diet proceedings for consideration.

7.7 Netherlands

7.8 New Zealand

The Privacy Act 1993 sets out how agencies may collect, store, use and disclose personal information. One of the binding principles relates to the purpose of collection and use of personal information. Personal information must not be collected unless it is collected for a lawful purpose and it is necessary to collect the information for that purpose. In addition, personal information obtained in connection with one purpose must not be used for another (with exceptions).

The Official Information Act 1992 sets out release of information held by Government agencies. There are provisions within this Act for declining information such as protecting the personal privacy, put health and safety, commercial confidentiality and other reasons.

7.9 Singapore

[A]. The Personal Data Protection Act 2012 (PDPA) [32]

In Singapore, the *Personal Data Protection Act* 2012 (PDPA) establishes a data protection law that comprises various rules governing the collection, use, disclosure and care of personal data. It recognises the rights of individuals to protect their personal data, including rights of access and correction, and the needs of organisations to collect, use or disclose personal data for legitimate and reasonable purposes. The PDPA covers personal data stored in electronic and non-electronic forms. The data protection provisions in the PDPA generally do not apply to any public agency or an organisation in the course of acting on behalf of a public agency in relation to the collection, use or disclosure of the personal data. Nevertheless, public agencies are governed by the government policy on Data management, which is largely aligned with the requirements of the PDPA.

[B]. Transfer of personal data outside of Singapore[33]

The PDPA limits the ability of an organisation to transfer personal data outside Singapore. Transfer of personal data outside of Singapore needs written permission from the Personal Data Protection Commission which consists of not fewer than 3 but not more than 17 members appointed by the Minister. In essence, an organisation may transfer personal data overseas if:

(i) it has taken appropriate steps to ensure that it will comply with the Data Protection Provisions in respect of the transferred personal data while such personal data remains in its possession or under its control; and
(ii) that the oversea recipient is bound by legally enforceable obligations to provide to the personal data transferred a standard of protection that is comparable to that under the PDPA.

An organisation transferring personal data overseas is taken to have satisfied the requirements if:

(i) the individual whose personal data is to be transferred gives his consent to the transfer of his personal data;
(ii) the transfer is necessary for the performance of a contract between the organisation and the individual, or to do anything at the individual's request with a view to his entering a contract with the organisation;
(iii) the transfer is necessary for the conclusion or performance of a contract between the organisation and a third party which is entered into at the individual's request, or which a reasonable person would consider to be in the individual's interest;
(iv) the transfer is necessary in certain situations where the consent of the individual is not required under the PDPA, such as use or disclosure necessary to respond to an

[32] Extracted from the Personal Data Protection Commission Singapore (https://www.pdpc.gov.sg/legislation-and-guidelines/overview)
[33] Extracted from the Personal Data Protection Commission Singapore, The Transfer Limitation Obligation (Chapter 19) (https://www.pdpc.gov.sg/docs/default-source/advisory-guidelines/the-transfer-limitation-obligation-(chapter-19).pdf?sfvrsn=0)

emergency that threatens the life, health or safety of an individual. In such cases, the organisation may only transfer personal data if it has taken reasonable steps to ensure that the personal data will not be used or disclosed by the recipient for any other purpose;

(v) the personal data is data in transit; or

(vi) the personal data is publicly available in Singapore.

Nevertheless, the proposed data shared must be anonymised so that they cannot be individually identifiable, otherwise consent must be sought from the individuals for their data to be shared, which will be impractical. In addition, there should be Memorandum of Understanding amongst the ICMRA sub-group members regarding the sharing and protection of the data.

7.10 South Africa

All cases received at the National Centre are kept confidential and information is not shared with any third party.

7.11 Sweden

Patient data submitted to the MPA is handled and kept confidential according to the national law.

7.12 Switzerland

The legal bases applicable in Switzerland to the disclosure of data by government administrative authorities and the sharing of data with foreign partner authorities are as follows:

- European Convention on Human Rights (ECHR, SR 0.101)
- Federal Constitution of the Swiss Confederation (Cst, SR 101)
- Federal Act on Data Protection (Data Protection Act, FADP, SR 235.1)
- Ordinance to the Federal Act on Data Protection (Data Protection Ordinance, DPO, SR 235.11)
- Specifically relating to therapeutic products legislation: Federal Act on Medicinal Products and Medical Devices (Therapeutic Products Act, TPA, SR 812.21), Art. 61 ff.

Swissmedic, the Swiss Agency for Therapeutic Products, records safety signals associated with medicinal products, vaccines, labile blood products and veterinary medicines based on reports of adverse drug reactions (ADR) from within Switzerland. If the investigations confirm a new risk, Swissmedic initiates the necessary measures.

Swissmedic's tasks also include the assessment of international data on the safety of medicines.

The ADR reports submitted to Swissmedic usually contain a substantial amount of health-related information for the person concerned, and this information is often not anonymised. Such data are considered to be *sensitive personal data* under the terms of the Data Protection Act. Not

infrequently, the person concerned can be identified by combining the information about their health. This is all the more the case because the reports generally contain the initials of the people concerned – at least such is the intention of the agency's ADR report form. In order to prevent a person from being identified in this way, complete anonymisation of personal data is essential, otherwise such data may not be forwarded.

In theory, it would be conceivable for the person concerned in each case to be asked to agree to the disclosure of the sensitive data in accordance with the Data Protection Act. However, since data will be shared on a huge scale (big data, i.e. possible range of several billion gigabytes), it does not appear realistic or practical to obtain such individual consents.

In the course of the ordinary revision of the TPA, a legal basis was created that will enable the Federal Council in future to enter into international agreements on the *Disclosure of confidential data, including personal data,* but not sensitive personal data, to foreign authorities or international organisations, where this is required for the implementation of the TPA. However, it is not clear when this new provision will enter into force. There is no appropriate legal basis at present.

The existing Memorandums of Understanding and international treaties do not currently allow ADR reports containing personal data to be forwarded to third countries in terms of a "big data transfer".

By contrast, *non-confidential data*, i.e. data containing no personal information, can today be disclosed to the competent foreign authorities or international organisations.

To sum up, it can be stated that the legal situation in Switzerland currently allows only a limited sharing of ADR data.

8 Opportunities for collaboration on “Big Data & Analytics” among ICMRA Pharmacovigilance Sub-Group Members

8.1 Australia

8.1.1 Challenges/Gaps

The TGA is interested in working with other regulators to identify common challenges and gaps.

8.1.2 Sharing Data

The TGA contributes spontaneous adverse event reports to the WHO-UMC for inclusion in Vigibase.

8.1.3 Sharing Capacity (Tools and Expertise)

While the use of Big Data techniques for pharmacovigilance is in the very early stages, the TGA is very interested in the experiences of other regulators, and keen to participate in knowledge-sharing activities.

8.2 Brazil

8.2.1 Challenges/Gaps

Brazil is interested in working with other regulators to identify common challenges and gaps

8.2.2 Sharing data

Brazil participates in the WHO International Drug Monitoring Program for adverse reactions data.

8.2.3 Sharing capacity (tools and expertise)

As already mentioned, in Brazil, there is an expectation to acquire a system which is ready to share automatically to others systems. This possibility will enhance our capacity to share.

Currently, our system is not designed to share, because there is no interoperability with other systems.

8.3 Canada

8.3.1 Challenges/Gaps

Health Canada is interested in collaborating with other regulators to improve and enhance the post-market pharmacovigilance for all health products. However, a number of challenges, as stated previously, need to be addressed. Particularly, a definition of "big data" within the context of pharmacovigilance needs to be established and harmonized. In addition, the integration and utilization of big data into the current surveillance frameworks requires clearly defined objectives and plans, including financial and human resources for development, implementation, validation and maintenance of an infrastructure.

8.3.2 Sharing data

Health Canada participates in the WHO International Drug Monitoring Program for adverse reactions data.

8.3.3 Sharing capacity (tools and expertise)

With respect to health care data sharing, this is hindered by privacy and other laws across Canada and internationally. However, other approaches of data sharing maybe explored without the need of having the "raw data." A number of approaches are available in the ENCePP Guide on Methodological Standard in Pharmacoepidemiology. [34]

[34] http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml

Since the benefits of big data in post-market surveillance are increasingly being recognized, Health Canada is eager to work with other regulatory agencies on this project.

8.4 Ireland

8.5 Italy

8.5.1 Challenges/Gaps

AIFA does not currently have any involvement in data analytics projects for non-conventional data sources, such as social media or patient forums, however, there is great interest in learning from other agencies experiences in this area. AIFA took part to the IMI Innovative Medicines Initiative 9th Call with the LetUsTalk proposal and to a call Horizon for SME regarding a development of a semantic tool for monitoring of ADR in open data source and social media with a group from Modena University.

8.5.2 Sharing Data

Italy has been participating in the WHO International Drug Monitoring Program for adverse reactions data since 1975.

8.5.3 Sharing Capacity (Tools and Expertise)

AIFA is interested in the experiences of other regulators, and welcomes proposals in knowledge-sharing activities.

8.6 Japan

8.6.1 Challenges/Gaps

The PMDA is interested in working with other regulators to identify common challenges and gaps.

8.6.2 Sharing Data

The PMDA contributes spontaneous adverse drug reaction reports to Eudravigilance and the WHO-UMC for inclusion in Vigibase

8.6.3 Sharing Capacity (Tools and Expertise)

The PMDA is very interested in the experiences of other regulators, and am happy to participate in knowledge-sharing activities.

8.7 Netherlands

8.8 New Zealand

8.8.1 Challenges/Gaps

Medsafe is interested in working with other regulators to identify common challenges and gaps and how these maybe addressed.

8.8.2 Sharing Data

New Zealand participates in the WHO programme for International Drug Monitoring through the provision of adverse reactions data (Vigibase).

8.8.3 Sharing Capacity (Tools and Expertise)

New Zealand is keen to participate and work with other regulators.

8.9 Singapore

8.9.1 Challenges/Gaps

HSA is interested in working with other regulators to identify common challenges and gaps.

8.9.2 Sharing Data

HSA contributes spontaneous adverse event reports to the WHO-UMC for inclusion in VigiBase.

8.9.3 Sharing Capacity (Tools and Expertise)

While the use of Big Data techniques for pharmacovigilance is in the very early stages, HSA is very interested in the experiences of other regulators, and keen to participate in knowledge-sharing activities.

8.10 South Africa

8.10.1 Common Challenges/Gaps

South Africa is interested in working with other regulators in order to strengthen the vigilance within the country. Key gaps include development of data management infrastructure that allows for processing for various data types and sources. Capacity-building in coding, data analysis and reporting.

8.10.2 Sharing Data

South Africa participates in the WHO International Drug Monitoring Program for adverse reaction reporting which facilitates data sharing.

8.10.3 Sharing Capacity (Tools and Experience)

8.10.4 Additional Questions

1. Can regulators learn by **sharing different approaches to signal detection** in large data sets, including assessment of signals in different subpopulations and controlling for confounders? Yes – absolutely. Building analytical capacity within regulatory authorities, for Spontaneous reporting systems as well as other pharmacovigilance systems is urgently needed, particularly in resource limited settings.

2. What role can regulators have in **encouraging healthcare professionals** to enter adverse event/ outcome data into national electronic medical records or into registries? Clinically

relevant and meaningful feedback is key as well as providing extremely easy and accessible systems for recording safety data. In addition, collecting and disseminating data from local active hospital-based surveillance studies for ADRs provides very important and useful information to hospital managers, pharmacy and therapeutics committees, medical and pharmacy schools of the important roles ADRs play in contributing to the national burden of morbidity and mortality, the extent to which they are preventable and which medicines are the most commonly implicated in the local context. Training of nurses in the detection of ADRs is also key – particularly in countries where nurse-led primary health care is the norm.

3. How can use of big data sets **complement existing spontaneous reporti**ng and epidemiological approaches? Detection of rare and long term

4. Does big data have a role to play in collecting **adverse event data following vaccination**, given increasing provision of vaccination services in other than general practitioner settings? Yes – epidemiological studies, ecological studies can be useful.

5. What methods could be developed to use big data to **identify adverse events that emerge over a longer period of time** (and may be missed in spontaneous reporting approaches) In South Africa, prospective pregnancy exposure registries (which include built-in control groups) and monitor large cohorts of pregnant women and their infants over time is one method that has been implemented. Birth defect surveillance is another approach being implemented. Long term disease cohorts (e.g. HIV and TB cohorts), and large-linked administrative data sets could assist depending on the quality of the data collected.

6. **Could one regulator ask another regulator to interrogate their big data set to enable a question on causality ques**tion to be asked e.g. through sequence symmetry analysis? (if development of common data models and actual sharing of big data sets between major regulators is potentially some years away) The implications on sovereignty over such data and ethics would need to be discussed. Would EMA, Ireland, the Netherlands countries be willing to share their big data sets with South Africa? Which direction is it expected that the data will flow?

7. **Are there opportunities for sharing experiences in the use of large data sets – such as the value of different di**sproportionality methods and other statistical approaches for data mining in big data sets, and collaboration in developing common data models and potentially in linking data sets? This is worth exploring. There needs to be stronger collaborations between regulators and academia, such as departments of public health, biostatistics. WHO and similar agencies should offer data analysis training for countries

interested in using large data sets. A consortium of regulators (overseen by a body such as WHO) who meet regularly to develop such models would be useful. Before linking of data sets can be considered, there should be discussions on harmonisation of data dictionaries, coding etc so that data can be compared across sites and countries.

8.11 Sweden

8.11.1 Challenges/Gaps

The MPA is currently not involved in any "Big Data-project". There is however an interest to discuss with other concerned agencies how to develop activities and tools within this area for the monitoring of ADR and to perform quality signal detection on Big Data.

8.11.2 Sharing Data

Sweden has participated in the WHO International Drug Monitoring Program for adverse reaction data for many years now. Sweden also sends ICSRs to EudraVigilance according to applicable regulations. As long as data is fully anonymised there is currently possible to share data with other competent agencies.

8.11.3 Sharing Capacity (Tools and Expertise)

The Swedish MPA is interested in a work-sharing activity to take part in and discuss proposals related to the development of tools and expertise in Big Data.

8.12 Switzerland

8.12.1 Common Challenges/Gaps

Biggest challenge: Security and privacy legal considerations for sharing of data between countries

8.12.2 Sharing Data

Swissmedic contributes with spontaneous adverse event reports to the WHO-UMC for inclusion in Vigibase. Switzerland has been participating in the WHO International Drug Monitoring Program for adverse reactions for many years now.

8.12.3 Sharing Capacity (Tools and Expertise)

While the use of Big Data techniques for pharmacovigilance is in the very early stages, Swissmedic is interested in the experiences of other regulators, and keen to participate in knowledge-sharing activities.

8.13 UK

8.13.1 Challenges/Gaps

The MHRA is interested in working with other regulators to identify common challenges and gaps.

8.13.2 Sharing Data

The MHRA contributes spontaneous adverse drug reaction reports to Eudravigilance and the WHO-UMC for inclusion in Vigibase.

8.13.3 Sharing Capacity (Tools and Expertise)

While the use of Big Data techniques for pharmacovigilance is in the very early stages, the MHRA is very interested in the experiences of other regulators, and keen to participate in knowledge-sharing activities.

9 Common Challenges

The common challenges identified include:

Security and privacy

- Privacy & legal considerations for sharing of data between countries

Partnerships

- Building domestic and international partnerships to access data sources that are not held by the regulator.

Infrastructure & capacity

- Building internal capacity to leverage new data sources, including both IT infrastructure and the expertise to use it effectively.
- Building infrastructure and common data models to support datasets and develop innovative methods to link together different data sources.
- Resources required to develop, evaluate and use data may be substantial and will involve multi-disciplinary teams.

Standards

- Developing international standards, principles and best practices for sharing, combining and validating data from multiple sources.
- Establishing international harmonized standard to accept results of big data analysis by other regulatory agencies.

- Developing common data models to share data.
- Different coding systems for spontaneous AE databases and health care data, and requires extraction and mapping of ADRs.

Data analysis

- Establishing methods and standards to evaluate data for secondary uses, which may be biased due to a number of factors, including practitioner behaviour, administrative dataset is mainly for financial purposes, EMR is primary for clinical management.
- Incomplete data due to partial participation by healthcare providers (public vs. private).
- Challenges in converting free text data to structured data.
- The rules governing data collection may change over time, and electronic access to historical records may be limited.

10 Proposed approaches

Despite of all the common challenges, all sub-group members have expressed an interest in collaborating further on big data initiatives for pharmacovigilance. However, a more focussed work- sharing is required for the benefit of this sub-working group because of the following:

- All sub-group members are already participating in the World Health Organization-Uppsala Monitoring Centre (WHO UMC) i.e all are submitting Individual Case Safety Reports (ICSRs) to the VigiBase. Therefore, sharing ICSRs within sub-group members will be duplication of work since these data are already available in the VigiBase.
- For health care data, data sharing will be hindered by privacy and other laws within and across each member jurisdictions. In addition, other methods are required to assist with the unique issues that arise from working with non–ICSR data.
- It is noted that other ICMRA members in addition to those that contributed to this paper, including the US FDA and the EMA, have various pharmacovigilance big data initiatives at various stages of maturity. Current regulatory initiatives and experiences regarding validation studies for the use of real-world data in pharmacovigilance of all ICMRA members could be shared, to minimize redundancy, and the best practices and validation findings of these initiatives could be shared when they become available.
- The integration and utilization of non-conventional/unstructured data sources in pharmacovigilance is also resource-intensive and may be potentially redundant. An established framework is required for knowledge transfer of the results of existing research initiatives to validate these data sources for use in pharmacovigilance. For example, issues surrounding such data can be discussed among all sub-group members based on analytical results shared by an agency.

Based on the sub-group members' contributions in Section 6, 7, and 8, a two-stage approach for working sharing maybe considered.

- **Short-term:** All sub-group members have expressed an interest in a more focussed knowledge sharing on a technical level to identify specific challenges and gaps. In particular, sub-group members raised the needs of best practices for utilizing real-world data sources such as EHRs, EMRs and AHD together with traditional SRS data. Ensuring that aggregated data from each member are consistent and comparable is a prerequisite to more focussed research collaborations. The development of guidance or technical standards, if this is feasible, would be beneficial to all sub-group members.
- **Long-term:** Collaborative development of study protocols, including: A common observational study protocol on a project of common interests for each member to conduct within their databases; and a meta-analysis study protocol with common data extraction procedure for each member's databases. These approaches will facilitate collaboration and will overcome the privacy and security issues.

11 Additional Questions

The following are additional questions may be of interest for this sub-working group.

8. Can regulators learn by **sharing different approaches to signal detection** in large data sets, including assessment of signals in different subpopulations and controlling for confounders?
9. What role can regulators have in **encouraging healthcare professionals** to enter adverse event/ outcome data into national electronic medical records or into registries?
10. How can use of big data sets **complement** existing spontaneous reporting and epidemiological approaches?
11. Does big data have a role to play in collecting **adverse event data following vaccination**, given increasing provision of vaccination services in other than general practitioner settings?
12. What methods could be developed to use big data to **identify adverse events that emerge over a longer period of time** (and may be missed in spontaneous reporting approaches)
13. **Could one regulator ask another regulator to interrogate their big data set** to enable a question on causality question to be asked e.g. through sequence symmetry analysis? (if development of common data models and actual sharing of big data sets between major regulators is potentially some years away)
14. Are there **opportunities for sharing experiences in the use of large data sets** – such as the value of different disproportionality methods and other statistical approaches for data mining in big data sets, and collaboration in developing common data models and potentially in linking data sets?

12 Summary and Conclusion

Pharmacovigilance is a critical and essential component for strengthening health products safety. With the current limitations of SRSs, all regulatory agencies are forced to find new ways to enhance the quality and quantity of ADRs, including the use of big data in pharmacovigilance. All sub-group members agreed to share more experiences and knowledge in utilizing big data for regulatory purpose as a first step. Such activity can further identify common challenges and gaps for regulatory harmonization and collaboration on use of big data in pharmacovigilance.

While each member acknowledges benefits of big data in post-market surveillance, a number of challenges were highlighted (Section 9), particularly with the integration and utilization of big data into the current pharmacovigilance framework. Despite of all the common challenges, all sub-group members have expressed an interest in collaborating further on big data initiatives for pharmacovigilance. However, a more focussed work- sharing is required for the benefit of this sub-working group because of the following:

- All sub-group members are already participating in the World Health Organization-Uppsala Monitoring Centre (WHO UMC) i.e all are submitting Individual Case Safety Reports (ICSRs) to the VigiBase. Therefore, sharing ICSRs within sub-group members will be duplication of work since these data are already available in the VigiBase.
- For health care data, data sharing will be hindered by privacy and other laws within and across each member jurisdictions. In addition, other methods are required to assist with the unique issues that arise from working with non–ICSR data.
- It is noted that other ICMRA members in addition to those that contributed to this paper, including the US FDA and the EMA, have various pharmacovigilance big data initiatives at various stages of maturity. Current research initiatives and validation studies for the use of real-world data in pharmacovigilance of all ICMRA members could be shared, to minimize redundancy, and the best practices and validation findings of these initiatives could be shared when they become available.
- The integration and utilization of non-conventional/unstructured data sources in pharmacovigilance is also resource-intensive and may be potentially redundant. An established framework is required for knowledge transfer of the results of existing research initiatives to validate these data sources for use in pharmacovigilance. For example, issues surrounding such data can be discussed among all sub-group members based on analytical results shared by an agency.

Based on the sub-group members' contributions in Section 6, 7, and 8, a two-stage approach for working sharing maybe considered.

- **Short-term:** All sub-group members have expressed an interest in a more focussed knowledge sharing on a technical level to identify specific challenges and gaps. In particular, sub-group members raised the needs of best practices for utilizing real-world data sources such as EHRs, EMRs and AHD together with traditional SRS data, either in

a single database or in a distributed method. Ensuring that aggregated data from each member are consistent and comparable is a prerequisite to more focussed research collaborations. The development of guidance or technical standards, if this is feasible, would be beneficial to all sub-group members.

- **Long-term:** Collaborative development of study protocols, including: A common observational study protocol on a project of common interests for each member to conduct within their databases; and a meta-analysis study protocol with common data extraction procedure for each member's databases. These approaches will facilitate collaboration and will overcome the privacy and security issues.

In addition, the following recommendations were made from the sub-group members to the ICMRA Management Committee to promote additional collaboration:

- A standing working group should be formed to facilitate knowledge transfer between ICMRA members with respect to leveraging big data analytics for pharmacovigilance.
- All ICMRA members should be invited to notify partners of on-going research initiatives and validation studies in using non-conventional data sources to complement existing pharmacovigilance activities, to minimize redundancy.
- All ICMRA members should be invited to share the results of research initiatives and validation studies in using big data sources to complement existing pharmacovigialnce activities, when they become available.
- Best practices, including coding standards, for combining multiple sources of real-world data, including EHRs, EMRs and AHD with traditional SRS data, should be shared.
- Develop a common observational study protocol on a project of common interests for each member to conduct within their databases. Findings are then analyzed and shared within members.
- Develop a meta-analysis study protocol with common data extraction procedure for each member's databases.