Reflections on the Findings and Conclusions of the European Medicines Agency (EMA) Workshop on “Big Data” and Healthcare

ADEPT 4 Workshop
Dr Alison Cave,
Principal Scientific Administrator,
Pharmacovigilance and Epidemiology Department
Disclaimer

The views expressed in this presentation are my personal views and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medicines Agency or one of its committees or working parties.
Objectives

1. Why now?
2. Key Messages
3. New Regulatory Initiative
4. Conclusions
Objectives

1 Why Now?
An increasing number of medicines with genomic mechanism of action and/or genomic biomarkers enabling smaller, focused RCTs but creates other challenges.
Genomic Based Mechanism of Action

• Cystic fibrosis is caused by one of nearly 2000 mutations.
• CF drug, ivacaftor which targets $G551D$ mutation in the $CFTR$ gene (4% of CF population).
• Delivers increases in $FEV_1 \sim 10\%$.

**Indication gradually expanded to covers further mutations**

**The future**
Challenge of determining the level of evidence required to extend indications when further mutations are identified.
An increasing number of medicines with genomic mechanism and/or genomic biomarkers enabling smaller, focused RCTs but increasing uncertainties

Innovative medicines and personalised prescribing creates regulatory challenges.
But for other diseases the genetic risk is less predictable e.g. Alzheimer’s, Parkinson’s

How do you identify patients to be treated prophylactically and how do you assess the benefit-risk profile?
An increasing number of medicines with genomic mechanism and/or genomic biomarkers enabling smaller, focused RCTs but increasing uncertainty.

Innovative medicines and personalised prescribing creates regulatory challenges.

**Rare diseases may be associated with more limited information at authorisation**
**Strimvelis** - Corrective gene therapy for children with SCID-ADH (Severe Combined Immunodeficiency due to adenosine deaminase deficiency). Occurrence: 0.22-0.68 per 100,000 population

- 12-patient pivotal study; Open label
- Primary outcome: 3-year survival
- Secondary outcome: severe infections

- 3-year survival: 12/12
- 9/12 successful response
- 12/18 auto-immune AEs

**Uncertainties**

- Long term durability of benefit (comparison with stem cell transplant)
- Late failure – need for further treatment eg stem cell transplant
- Late toxicity
- Long-term immunogenicity

Conditional MA
Number of applications requesting conditional marketing authorisation at submission, by year of submission

107 post-authorisation obligations
(of these, 57 obligations were fulfilled before June 2016)

Categories of specific obligations imposed to companies
- Final results from clinical studies or pool of studies
- Interim results of clinical trials
- Additional analysis
- Quality data
- Other measures

How timely was the submission of specific obligation results?
- Due early +/- 1 month: 33
- Early (1-6 months): 15
- >1 year early: 4
- Late (6-12 months): 1
- Late (>12 months): 2

>90% of completed specific obligations did not have major changes to their scope
≈70% of specific obligations were completed within specified timelines
An increasing number of medicines with genomic mechanism and/or genomic biomarkers enabling smaller, focused RCTs but increases uncertainties.

Innovative medicines and personalised prescribing creates regulatory challenges.

Rare diseases to may be associated with more limited information at authorisation.

Unknown generalisability of RCT results to normal clinical practice: Need for new approaches to gather complementary evidence.
Happich et al (ISPOR 19th Annual European Congress, GETREAL) developed a propensity score model that predicts participation in either a RCT (JMDB) or the real world (FRAME), given a set of common total baseline characteristics.

Resulting propensity scores were used to assess the overlap between the two cohorts.
An increasing number of medicines with genomic mechanism and/or genomic biomarkers enabling smaller, focused RCTs but increases uncertainty.

Innovative medicines and personalised prescribing creates regulatory challenges.

Rare diseases to may be associated with more limited information at authorisation.

Unknown generalisability of RCT results to normal clinical practice: need for new approaches to gather complementary evidence.

Additional data sources are needed to better monitor risk/benefit in high risk groups often excluded from clinical trials.
Meeting needs? Pharmacoepidemiological paediatric safety studies

2006 Regulation (EC) 1901/2006 on medicinal products for paediatric use
Main provisions applied from July 2008 & January 2009
PDCO established

2009-2013: <1% of P’epi safety studies conducted in paediatric populations
Exclusion of Elderly People from Randomized Clinical Trials of Drugs for Ischemic Heart Disease

Florence T. Bourgeois, MD, MPH,†† Liat Orenstein, MSc,‡ Sarita Ballakur,‡ Kenneth D. Mandl, MD, MPH,§‡ and John P. A. Ioannidis, MD, DSc†‡

OBJECTIVES: To measure exclusion of elderly adults from randomized trials studying drug interventions for ischemic heart disease (IHD) and describe the characteristics of these trials.

SETTING: Cross-sectional analysis.

SETTING: Interventional clinical trials studying a drug intervention for IHD that started in 2006 and after were identified in ClinicalTrials.gov. Data were extracted on study features, including age-based inclusion criteria. Data on participants and their age distribution were collected from trial publications, investigator inquiry, and result data in ClinicalTrials.gov.

PARTICIPANTS: Individuals aged 65 and older.

MEASUREMENTS: Proportion of trials excluding individuals based on age, mean age of trial participants, and proportion of enrolled participants aged 65 and older 75 and older.

RESULTS: Of 839 identified trials, 446 (53%) explicitly excluded elderly adults. The most frequent upper age limits were 80 (n = 164) and 75 (n = 114), with a median upper age limit of 80 (interquartile range 75–80). Trials with upper age limit exclusions tended to be smaller (median number of participants 100 vs 201, P < .001) and were more likely to be funded primarily by nonindustry sources (75.3% vs 70.0%, P = .006). The overall mean age of trial participants was 62.7 (mean maximum age 74). The estimated proportion of participants aged 65 and

CONCLUSION: Despite the high burden of IHD in elderly adults, the majority of drug trials do not enroll participants reflective of age-related prevalence of the disease. J Am Geriatr Soc 2017.

Key words: ischemic heart disease; evidence-based medicine; research methodology

Of 839 identified trials, 446 (53%) explicitly excluded elderly adults.
An increasing number of medicines with genomic mechanism and/or genomic biomarkers enabling smaller, focused RCTs but increases uncertainty.

New innovative medicines and personalised prescribing creates regulatory challenges.

Welcome activity in the rare disease area to meet unmet medical needs is associated with more limited information at authorisation.

The high internal validity of clinical trials at the expense of external validity demands new approaches to gather complementary evidence.

Additional data sources are needed to appropriately monitor risk/benefit in high risk groups often excluded from clinical trials.

Increasing interest in combination therapies to treat complex diseases creates regulatory challenges.
Ceftazidime-avibactam: a novel cephalosporin/β-lactamase inhibitor

Clinical Pharmacist | 10 MAY 2017 | By Sharanie V. Sims, Elizabeth A. Neuner, Robert A. Bonomo

Sulphonylurea compared to DPP-4 inhibitors in combination with metformin carries increased risk of severe hypoglycemia, cardiovascular events, and all-cause mortality

Jan W. Eriksson, Johan Bodegård, David Nathanson, Marcus Thuresson, Thomas Nyström, Anna Norhammar

RHEUMATOID ARTHRITIS
Comparing durability of combination therapies
According to observations from a follow-up study of the RACAT trial looking at patients with rheumatoid arthritis who have suboptimal responses to methotrexate, triple therapy with methotrexate, sulfasalazine and hydroxychloroquine is more durable than combined methotrexate–etanercept therapy. Of the 289 patients followed up, 78% remained on triple therapy at 1 year compared with 63% who remained on methotrexate–etanercept therapy; significantly more patients changed from methotrexate–etanercept therapy to triple therapy than vice versa (P = 0.005).

• Define the Big Data landscape from a regulatory perspective

• Clarify the opportunities and the challenges

• Identify what is needed for Big Data to be exploited to support medicines development and regulatory decision making
Objectives

1. Regulatory Challenges
2. Key Messages
Defining the Big Data Landscape
Big Data – how will we meet it?

Big data

Data sets that are so large or complex that traditional data processing applications are inadequate (Wikipedia)

In the context of Medicines Regulation it could mean data in large amounts or of a complex nature reaching regulatory authorities in the margins of the more traditional analysed and structured data.

Data lying underneath the regulatory submissions, for which it would be crucial to understand their presence and the robustness by which they were generated in order to make a competent evaluation of the submission as a whole.
The data landscape: which data?

Datasources:
- Clinical databases (Prescriptions, EHRs and registries)
- Social media data/m-health
- Clinical trial data
- Imaging data
- ‘Omic data (Genetic, proteomic, metabolomic)
- Published literature
- Regulatory data (ADR, sales, Safety updates, PASS, PAES)
90%
Of the world’s data has been created in the past 2 years.

24 months
Frequency at which electronic healthcare data doubles

75%+
Percentage of patients expected to use digital health services in the future
Data per individual

70% Social and Environment And Behavioral

20% Genomics Factors

10% Clinical Factors

1100 Terabytes Generated Per lifetime

6 Terabytes Per lifetime

0.4 Terabytes Per lifetime
Defining the Big Data Landscape

Data Accessibility and Integration
Sentinel is a network of distributed data approach which allows the FDA to rapidly and securely access information via a CDM from large amounts of electronic healthcare data, such as EHRs, insurance claims data and registries. Pilot project delivers access to 99 million patient lives, 2.9 billion drug prescriptions and 38 million acute hospital stays.

The CNODES network delivers access to the health and prescription records of over 40 million people and a widely distributed network of academic and data analytics experts to rapidly evaluate the risk:benefit profiles of medicines.

OHDSI is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. All the solutions are open-source. Currently the community has converted >50 databases covering >660 million patients.
Bringing the data together is very hard. It needs to be “standardised”, structured and stored together to deliver insight.

Data is siloed at individual centres, hard to access, analyse and use.

Productivity tools (especially IT) built for individual local usage focusing on local data analytics solutions.

Data needs to be **FAIR**: Findable, Accessible, Interoperable and Reusable.

We need centralised IT solutions to store data safely and securely and enable machine learning solutions.
Regulating the internet giants

The world’s most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules
Defining the Big Data Landscape

Data Accessibility and Integration

Clinical Utility
Data Collection should be targeted

Clinical utility starts by asking the right question

Data should be collected to respond to questions that will translate to benefit

As a [job title]

I need [big data insights]

So that I can [make a decision my job expects me to]
• 15% of EMA evaluated medicines containing PGx information
  • Therapeutic indication (3.5%)
  • Posology and method of administration (4.4%)
  • Contraindications (6.4%)
Clinical Evidence Supporting Pharmacogenomic Biomarker Testing Provided in US Food and Drug Administration Drug Labels

Bo Wang, PharmD; William J. Canestaro, MSc; Niteesh K. Choudhry, MD, PhD

- 119 drug-biomarker combinations
- 43 (36.1%) had convincing clinical validity evidence
- 18 (15.1%) evidence of clinical utility
- 61 labels (51.3%) – clinical decisions based on results of biomarker test: 36 (30%) contained convincing clinical utility data

“It may be premature to include biomarker testing recommendations in drug labels when convincing data that link testing to patient outcomes do not exist.”

Sir Munir Pirmohammed, Nov 2016
Defining the Big Data Landscape

Data Accessibility and Integration

Clinical Utility

Differentiating causality from co-incidence
August 2010: “the use of oral bisphosphonates was not significantly associated with incident esophageal or gastric cancer”

Sept 2010: “we found a significantly increased risk of oesophageal cancer in people with previous prescriptions for oral bisphosphonates.”
Sources of Variability in Multiple Database Studies

SCCS >30 days
SCCS 15-30 days
SCCS 7-14 days
SCCS - 0-7 days
CXO - 30 days
CXO - 14 days
CC - Definite
NCC - Definite/Probable
NCC - Definite
Cohort - Antibiotics Definite/Probable
Cohort - Antibiotics Definite
Cohort - Definite/Probable
Cohort - Definite

Relative Risk
(Log Scale)

0.25 1 4 16

BIFAP
CPRD
Clinformatics
Mondriaan-UPOD
CPRD-R

PROTECT
Antibiotics and the risk of acute liver injury
Joint development of Common protocol
Independent conduct in different databases

Pharmacoepidemiology and Drug Safety
2016;156-165. DOI: 10.1002/pds.3968

SCCS: self-controlled case series, CXO: case cross-over, CC: case–control, NCC: nested case–control
Sources of Variability in Multiple Database Studies

- Consistent direction of effect estimate but of varying magnitude
- Study design should be a conscious decision
Sources of Variability in Multiple Database Studies

- Stringency and accuracy of definition increased strength of association
- Less stringency led to more false positives
- Outcome needs to be carefully defined.
Sources of Variability in Multiple Database Studies

**Study design**
- SCCS >30 days
- SCCS 15-30 days
- SCCS 7-14 days
- SCCS 0-7 days
- CXO - 30 days
- CXO - 14 days
- CC - Definite
- NCC - Definite/Probable
- NCC - Definite

**Outcome**
- BIFAP
- CPRD
- Clinformatics
- Mondriaan-UPOD
- CPRD-R

- Time window of exposure had substantial impact
- Careful definition of exposure window is essential
Sources of Variability in Multiple Database Studies

Study design
- SCCS >30 days
- SCCS 15-30 days
- SCCS 7-14 days
- SCCS 0-7 days
- CXO - 30 days
- CXO - 14 days
- CC - Definite
- NCC - Definite/Probable
- NCC - Definite
- Cohort - Antibiotics Definite/Probable
- Cohort - Antibiotics Definite
- Cohort - Definite/Probable
- Cohort - Definite

Outcome
- Disease stratification
- Comorbidities/medications
- Adherence
- Methodology for matching
Sources of Variability in Multiple Database Studies

Databases vary in the lifestyle factors recorded and the quality of their measurement making comparisons difficult.
Sources of Variability in Multiple Database Studies

Study design
- SCCS >30 days
- SCCS 15-30 days
- SCCS 7-14 days
- SCCS 0-7 days

Outcome
- CXO - 30 days
- CXO - 14 days
- CC - Definite
- NCC - Definite/Probable
- NCC - Definite

Exposure
- Cohort - Antibiotics Definite/Probable
- Cohort - Antibiotics Definite
- Cohort - Definite/Probable
- Cohort - Definite

Confounding adjustment
- Accuracy and completeness data across different parameters is variable
- Systematic evaluation of strengths and limitations is essential

Database
- BIFAP
- CPRD
- Cliniformatics
- Mondriaan-UPOD
- CPRD-R

Study population

Relative Risk (Log Scale)
Regulatory Challenges

Structured data (RCT) generated in accordance with strict guidelines and known provenance

High certainty

Unstructured, unvalidated data of unknown provenance

more uncertainty
Visualization of the topology of complex data from the U-BIOPRED consortium of adult severe asthma cohorts

Cohorts are generated following the integration of multiple biomarkers

- How are the individual components validated?
- How reproducible are the cohorts?
- How is data weighted within the algorithms to define the cohorts?
- How do you identify the stability of the cohorts over time?
- Are the cohorts translatable to a defined patient population?

Aim: how do we generate certainty for regulatory decision making?
Dementia prevention, intervention, and care

Prof Gill Livingston, MD, Andrew Sommerlad, MSc, Vasiliki Orgeta, PhD, Sergi G Costafreda, PhD, Jonathan Huntley, PhD, Prof David Ames, MD, Prof Clive Ballard, MD, Prof Sube Banerjee, MD, Prof Alistair Burns, MD, Prof Jiska Cohen-Mansfield, PhD, Claudia Cooper, PhD, Prof Nick Fox, MD, Laura N Gitlin, PhD, Prof Robert Howard, MD, Prof Helen C Kales, MD, Prof Eric B Larson, MD, Prof Karen Ritchie, PhD, Prof Kenneth Rockwood, MD, Elizabeth L Sampson, MD, Quincy Samus, PhD, Prof Lon S Schneider, MD, Prof Geir Selbæk, PhD, Prof Linda Teri, PhD, Naheed Mukadam, MSc

Published: 19 July 2017

DOI: http://dx.doi.org/10.1016/S0140-6736(17)31363-6

Nine lifestyle changes can reduce dementia risk, study says

By Fergus Walsh
Medical correspondent

© 20 July 2017 | Health | 82

Top Stories

High risk of 'unprecedented' UK downpours
A new analysis suggests there's a greater chance of the heavy rain that led to extensive flooding in 2014.
© 1 hour ago

Trump son-in-law denies Russia collusion
© 8 minutes ago
Defining the Big Data Landscape

Data Accessibility and Integration

Clinical Utility

Differentiating causality from co-incidence

Solutions
Changes in the Traditional Regulatory Paradigm

Currently

- Structured data (RCT) generated in accordance with strict guidelines and known provenance
- High certainty

Challenge

- Unstructured, unvalidated data of unknown provenance
- Turning data into knowledge
- More uncertainty

Need to develop a deep understanding of the data, to define the strengths and limitations so that the evidence arising from its analysis can be appropriately challenged

Solution
Interoperability and Harmonisation
Common data models
Minimal Data sets
Standards

Documenting the Strengths and Limitations enabling robust validation

Addressing privacy and Governance

Solutions

Accessibility
Data for the Common Good
Objectives

1. Why now?
2. Key Messages
3. New Regulatory Initiative
Mandate HMA / EMA Joint Task Force Big Data
Priority: Reinforce the scientific and regulatory capacity and capability of the network, Innovation and access to new medicines, Optimisation of the regulatory operations

Chair: Thomas Senderovitz, DK
Co Chair: Alison Cave, EMA

Members: DE, DK, ES, FI, HU, IE, NL, NO, RO, UK
The Task Force should **characterise** relevant sources of big data and define the main format, in which they can be expected to exist in

Identify areas of **usability and applicability** of data

**Gap analysis** – describe the current status of expertise, future needs and challenges

The Task Force will generate a **list of recommendations** and **Big Data Roadmap**
Objectives

1. Why now?
2. Key Messages
3. New Regulatory Initiative
4. Conclusions
Conclusions

- Vast amounts of healthcare data are continually being generated, offering huge opportunities but making it impossible to keep pace with all the information.

- Harnessing of the potential of big data by researchers and regulators is hindered by the fact that it is often unstructured, noisy and inaccessible.

- Deciding which data to collect starts by asking the right questions about the benefits sought and problems faced.

- Access to data is a significant hurdle especially for observational data. Mechanisms to integrate the data to generate meaningful knowledge is needed.

- Validation that associations are real is key for data to support regulatory decision making.
“We are all drowning in a sea of data and starving for knowledge”

Nobel Lecture 2002

Sydney Brenner

Nobel Prize in Physiology or Medicine
Thank you
European Medicines Agency
30 Churchill Place London
E14 5EU

www.ema.europa.eu
info@ema.europa.eu

alison.cave@ema.europa.eu