30 January 2022
Data Analytics and Methods Task Force
EMA/18271/2022

Learnings Initiative for Optimal Use of Big Data for Regulatory Purpose

Meeting report

30 November 2021

Chairs: Peter Arlett, EMA; Jesper Kjaer, Danish Medicines Agency

Objectives

Regulators need to continuously assess big data in combination with novel methodologies to take decisions on individual products that need to be based on comprehensive, valid and reliable evidence. Use of big data has however introduced new challenges and uncertainties around the quality of an increasing volume of complex data, new analytical approaches, new processes and the need for additional expertise and guidance. The Big Data Task Force Report – Phase 2 emphasised that assessment of big data requires continuous optimisation of current standards and specifications to provide all stakeholders with rules and guidance supporting evidence generation.

Building upon the experience of the submission of real-world evidence for regulatory purpose and the conclusions of previous workshops on data standardisation, metadata and artificial intelligence the European Medicines Agency organised on 30 November 2021 a webinar aiming to:

- learn from current experience of using real-world data for regulatory purpose
- discuss important challenges related to optimal use of real-world data, including data relevance, submission processes and training needs
- discuss means to support effective consultation with stakeholders, including industry, the regulatory network, academia, healthcare professionals and patients.

The webinar was divided into three plenary sessions and five parallel breakout sessions. Reports from the breakout sessions were presented and discussed in Plenary session 2. The Agenda and presentations of the webinar have been posted on the EMA website.

This document summarises the presentations and recommendations expressed in the breakout sessions. These will be fed back into internal processes, stakeholder consultations, support for individual product submission and ultimately guidelines for applicants and the regulatory network.

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

Plenary session 1: Lessons learned from the submission of big data for regulatory purpose

Chair: Alar Irs, State Agency of Medicines, Estonia

1. In the context of recent studies on real-world evidence (RWE) used for regulatory decision-making, **Elisabeth Bakker** (Department of Clinical Pharmacy and Pharmacology, University Medical Center Groningen, The Netherlands) presented the preliminary results of an ongoing EMA study on RWE in marketing of which phase I "**Marketing authorization applications Made to the European Medicines Agency in 2018-2019: What was the contribution of RWE?**" has recently been published.[1] RWD/RWE included in centralised marketing authorisation applications (MAA) and extensions of indications (EoI) submitted in 2018-2019 was characterised and it was shown that RWD/RWE was used in 40% of MAAs (mainly post-authorisation) and in 18% of EoIs (mainly pre- or post-authorisation). The majority of products were antineoplastic and immunosuppressants (35% MAA and 42% EoI). When RWD/RWE was used pre-authorisation, these were mainly supporting studies looking at efficacy/ effectiveness. Post-authorisation RWE mainly concerned RMP Category 3 studies looking at safety.

It was concluded that there is already widespread use of RWE to support evaluation of MAAs and EOIs submitted to the EMA, but that a qualitative approach is needed to further describe how RWE contributed to the regulatory applications. This is currently done in Phase II of the study which aims to characterise the RWE used by applicants to support efficacy/effectiveness claims and disease epidemiology as part of MAAs and EoIs in 2018-2019, to analyse the contribution of RWE to the Committee for Medicinal Products for Human Use (CHMP) decision making on these MAAs and EoIs and to provide guidance to industry on the submission of RWE within these applications. Preliminary Phase II data show that of the 16 products where RWE was included pre-authorisation, 8 were authorised, 6 withdrawn and 2 refused. The 19 products with post-authorisation RWE were authorised. MAAs were: 14 standard, 4 conditional and 3 exceptional. The pre-authorisation RWE was mainly used for contextualisation of the efficacy data, e.g., for natural history or as an external comparator. Out of 16 dossiers including pre-authorisation RWE, in 4 the RWE was critical for decision making, of which two examples were presented. In these cases, the final appraisal of the RWE by the CHMP was positive, which was also in line with the scientific advice provided by the scientific advice working party (SAWP) in a procedure prior to submission of the MAAs during the drug development phase. In both cases, however, post-authorisation studies had to be performed as imposed by the CHMP.

Overall, an increasing use of RWD in applications has been shown and the effort of applicants to provide as high-quality as possible RWD in applications to address regulatory requirements are acknowledged. Nevertheless, there are still challenges for both applicants and regulators. In this light, there is work in progress to provide guidance to applicants on the use of RWE based on the acquired experience. Additionally, the results of the ongoing study should provide relevant information to be included in future guidance documents.

2. Jeremy Rassen (Aetion) presented the results of a study on the **role of real-world evidence in FDA-approved new drug and biologics license applications from 2019-2021**. This study

[1] Flynn R. Et al. Marketing Authorization Applications Made to the European Medicines Agency in 2018– 2019: What was the Contribution of Real- World Evidence? Clin Pharmacol Ther. 2022;111(1):90-7.

was published in January 2022.[2] This study asked several questions: How often is RWE used in FDA approvals? When used, what is RWE used for? How impactful are RWE studies in establishing evidence of safety and/or effectiveness? Do RWE studies appear in product labels? What feedback has FDA given on RWE study quality?

Overall, the authors found that 116 approvals incorporated RWE in any form. Of these approvals, 88 included an RWE study intended to provide evidence of safety (43), effectiveness (15) or both (30) and 83 included RWE to provide therapeutic context. Among the 88 approvals supporting safety or effectiveness, RWE for 65 influenced FDA's final decision; 9% provided substantial or primary evidence, 65% provided supportive evidence and 26% provided evidence that was not adequate or concerned RWE studies that FDA did not address. Specific FDA feed-back was received on RWE submitted (which included methodological issues, sample size concerns, omission of patient level data and other limitations) for 42% of the 88 approvals, and RWE was included in product label for 33% of these approvals. The 88 approvals spanned 18 therapeutic areas, predominantly oncology, neuroscience, infectious disease and endocrinology & metabolism. .FDA's feedback on RWE study quality.

Based on these findings, the presenter concluded that use of RWE is frequent and is included in the vast majority of US approvals. Successful use of RWE in regulatory approvals required fit-for-purpose data, good study design, appropriate data collection, thoughtful data analysis and proactive communication with FDA.

3. Talita Duarte-Salles (Real World Epidemiology Research Group, IDIAPJGol, Barcelona-Spain) presented **the Charybdis project – use of real-world data from multiple countries and opportunities for regulatory purpose**. This study aimed to address unanswered questions as regards the COVID-19 patient trajectory, such as: who gets tested, infected and hospitalised; what are their symptoms and outcomes; how different is COVID-19 from influenza. Target cohorts included patients with three different SARS-CoV-2 diagnostic criteria, patients hospitalised with a COVID-19 diagnosis or positive test and patients admitted in intensive services, and comparable cohorts with influenza. The data network included 13 databases in the US, 9 (from 6 countries) in Europe and 3 (2 countries) from Asia-Pacific. The network has published or is preparing 16 scientific manuscripts on various topics such as a comparison of COVID-19 and influenza, the characterisation of COVID-19 among population subgroups with comorbidities of interest (e.g., obesity, hypertension, cancer, etc), and the use of repurposed drugs in COVID-19 patients. The presenter concluded that the network was able to set up a global community to face a global problem, it highlighted the value of a common data model in RWD to provide comprehensive, valid and reliable evidence to regulatory authorities and that open collaboration requires full transparency.

4. Karin Van Baelen (Janssen, Chair of the EFPIA IEGU WG) presented the **pharmaceutical industry's experience and challenges with RWE submission in marketing authorisation applications.** As a starting point, the rationale to use RWE in marketing authorisations was emphasised:

- to contribute to medicines development, learning and regulatory decisions in nearly all phases and across different therapeutic areas and product characteristics;
- to support evaluation of marketing authorisation applications, extension of indications and adding of new populations, for example complementarily to clinical trials and more particularly in support to conditional marketing authorisations and approval of orphan medicines;

[2] Purpura C et al. The Role of Real- World Evidence in FDA- Approved New Drug and Biologics License Applications. Clin Pharmacol Ther. 2022;111(1):133-44.

- to address safety and effectiveness questions post-authorisation.
- RWE was also widely used in the context of COVID-19 for comparative effectiveness of COVID-19 vaccines and treatments and better understand COVID-19 high risk groups, temporal tends and vaccine effectiveness against emerging variants.

The following challenges of using RWE in different contexts were nevertheless highlighted:

- Advice procedure: an agile framework for advice procedures based on qualification opinions for registries and considerations of novel methodologies using artificial intelligence, RWE and digital tools would provide greater predictability to the outcome of RWE submisions.
- Guidelines: a suite of RWE guidelines, standards and framework for submissions and more clarity relating to GDPR/privacy are needed; the following guidelines are priorities for industry:
    - Data quality framework
    - Submission of patient level data (incl. format for submission and procedures for data submission and validation)
    - Submissions that include RWE (incl. design, choice of data source, etc)
    - Principles/best practices associated with analytical methodology
    - Optimised qualification procedure for registries.
- New technologies: a clear regulatory framework for evaluating new technologies in clinical development is needed.
- Transparency: there is limited transparency regarding RWE assessment within EPARs and, potentially, regarding decisions based on evidence from DARWIN EU.

Karin Van Baelen concluded that optimising how RWE can be used to support a wide range of regulatory decisions requires the ongoing collaboration of all those who are impacted by those decisions. The industry welcomes continued collaboration and interactive learning opportunities with all stakeholders involved.

Plenary session 2: Reports from the breakout sessions

Chair: Xavier Kurz, European Medicines Agency

1. Breakout session: Fitness of real-world data for regulatory purpose

Chair: Marjon Pasmooij, MEB; Rapporteur: Kelly Plueschke, EMA

After introductory presentations by John Concato, FDA, and Kelly Plueschke, EMA , the group addressed 4 questions.

i) What information about real-world data (RWD) should be included in regulatory applications to support decision-making?

- Everyone agreed that the first step of planning a study based on RWD is to clarify the research question(s). The second step is then to analyse the landscape of available date sources.
- A "feasibility analysis" as described in the CHMP Guideline on registry-based studies (EMA/426390/2021) can help identify which RWD source(s) is/are fit for purpose to answer the study question(s) taking into account a number of elements, e.g. the variables collected, care setting covered, frequency of data updates, quality management in place, as well as the limitations of the data sources.

- Dialogue with all relevant stakeholders including regulators, medicines developers and data custodians should be held as early as possible to identify possible hurdles and manage expectations.
- There is a need for transparency from study design through publication by data custodians of characteristics of their databases, as these are key to the feasibility analysis. The catalogues of data sources and studies currently being developed by EMA in the frame of the Big Data recommendations on Data discoverability and Data quality & representativeness (see the BDSG workplan) will contribute to increasing transparency.

ii) Is the feasibility analysis referred to in the CHMP Guideline on registry-based studies applicable to other RWD than registries?

- The group considered that the feasibility analysis should also cover other RWD such as electronic health care records, pharmacy data, but also digital health technologies.
- Members of the group highlighted the need to collect RWD on over-the-counter products as these are usually not captured in claims databases but are widely used and can provide evidence on real life use of such medicines.
- Patients reported outcomes were also considered important to consider and the group recommended the need for guidance on integration of patients' perspective into RWE, i.e. how patients'perspectives can be integrated as RWE in regulatory decision making.
- Members acknowledged the importance of informed consent, as well as the added value of research and data sharing.

iii) Should minimum quality requirements be established in submissions of RWD for regulatory purposes?

- The scope of "data quality" is large as it can relate to each steps of data processing (e.g. data collection, mapping, analysis etc.). Quality requirements should be context specific and will depend on the data elements needed and which impact each of them may have on study analysis/results to answer a specific research question.
- Definition and scope of data quality is crucial: does it cover data collection, mapping, analysis, or all of these? Practical difficulties of measuring and interpreting data quality were highlighted.
- Data source validation through validation studies was considered important to increase regulators' confidence in data quality, e.g. for cancer data. Members highlighted the importance of i) knowing how validation studies have been performed, ii) disseminating good examples of validation studies performed to verify the level of accuracy of RWD sources (e.g. linkage of cancer diagnosis in primary care with cancer registry) and iii) publishing results of validation studies for different data sources for benchmarking.

iv) What other important aspects of RWD to be addressed in submissions for regulatory purposes?

- Regulators and Health Technology Assessment bodies (HTA)/Healthcare payers are interested in the same data but for different purposes. It is therefore essential to include HTA needs as much and as early as possible during drug development (e.g. comparative data against standards of care) to save time and avoid duplication. This is being addressed through increased dialogues between relevant stakeholders through mechanisms such as parallel regulators/HTA scientific advice, regulators cluster meetings and joint multi-stakeholders fora.
- This continuous dialogue will help foster understanding of RWD and promoting convergence in its use for regulatory decision-making.

2. Breakout session: Platform for stakeholders' consultation

Chair Juan Garcia Burgos, EMA; Rapporteur: Carla Jonker, EMA

The objective of the breakout session was to provide examples of existing platforms for collaboration and to discuss which mechanisms or fora could be put in place to support stakeholders' consultations on content, format and processes of submission of real-world data for regulatory purposes. After introductory presentations by Elizabeth Vroom, Duchenne Parent Project, Álmath Spooner, AbbVie, and Marie-Helene Pinheiro, EMA, the three following aspects were addressed:

i) General remarks

- Stakeholders engagement is critical to allow progress in the area of real-world data; the scope of stakeholders to be involved should be broadened, e.g. data providers, health care systems and policy makers; there is particularly a need to resonate beyond the EU regulatory network and to reach society through awareness and communication campaigns, which foster public and institutional trust in the use of RWE.
- There is a need to reinforce the link with health care systems to facilitate use of data that they hold. It is also critical that a good balance between public health needs and commercial interests is achieved.
- Patient organisations have a key role in supporting the use of data and conveying a message that patients may trust and support the use of data.

ii) What is the experience with existing platforms, are there any gaps?

- Available platforms and mechanism engaging with patients, industry and health care professionals are working well but need to be made more agile and allow broader and more flexible participation of other stakeholders, for example those working in routine clinical practice; additional specific channels/tools for regular and more tailored communication with specific stakeholder groups, e.g. industry, should be considered.
- Interoperability between different European initiatives should be facilitated
- Learnings need to be shared, e.g. through research in specific disease area (e.g. pilots, research collaborations).

iii) For which aspects of use of RWE is it critical to consult stakeholders?

- Relevance of the real-world data and policy development on submission of individual patient level data, including data privacy aspects
- Collection of hospital medical data and patient relevant outcomes, taking example of what is being done in the field of rare diseases.
- Methodology for better stakeholder consultation and engagement

3. Breakout session: Process optimisation

Chair: Inka Heikkinen, EuropaBio; Rapporteur: Andrej Segec (EMA)

The objective of the breakout session was to identify what process optimisation is desirable and necessary in real-world data related processes for regulatory applications and decision-making. After introductory presentations by Helga Gardarsdottir, Utrecht University and Solange Corriol Rohou, Astra Zeneca, the breakout session discussed four questions:

i) At what stage should use of RWD be discussed with regulators, and with whom? Should there be different discussions for technical and regulatory questions?

Early discussion to allow for planning of pre- and post-licensing RWE generation was highlighted, and this can happen at existing business pipeline meetings, pre-submission meetings, etc, - before and during advice/assessment procedures. It was noted that clarity of proposals and consideration to existing guidance (eg ENCePP methodological guide and protocol checklist) was important to have an informed discussion. Other points included:

- Share learnings, training and education - work with stakeholders
- Discuss with regulators - via scientific advice, PRAC (PASS procedures) – or other fora
- RWE is sometimes discussed standalone, or as part of overall Scientific Advice package – if so, it needs to be given sufficient prominence (and be discussed in-depth)
- Proposals should aim to answer a well-defined research question (safety, effectiveness, effectiveness of RMMs) and consider whether the study, as proposed, is feasible.

Separation of technical and regulatory questions could be helpful if this streamlines and frees up resources (also noting existing triage of questions by EMA at pre-submission meeting phase). This could allow for separating specialised input required (eg regarding statistical methods) and would consider capacity for answers and bottlenecks. It was also noted that stakeholders should consider data output uses also more broadly (linking to HTA/PLEG). The importance of Q&As and best practices for clarity on regulatory/administrative questions was emphasized and could reduce the number of questions for discussion.

ii) The CHMP Guideline on registry-based studies recommends early discussions of proposals for use of registries in regulatory submissions. Should such recommendation be applied to other RWD sources? Should differences be made between data sources?

In the discussion, the breakout session noted the common use of claims data in the US for RWE vs use of registries. No reason was identified to distinguish between data sources in applying the recommendations from the registry-based guideline which were viewed as applicable more broadly to RWE sources.

Early engagement with details on the proposed methods is viewed as always beneficial as this allows more informed planning, with more details to follow during a procedure later (and pre-specified in a statistical analytic plan.

iii) How could other stakeholders than pharmaceutical companies contribute to process optimisation and what could be the vehicles through which such input could be provided?

The session noted the need for regulatory science research – need for evidence on processes (and improvements), which other stakeholders could support. Development of guidance and Q&As was noted again (see also in point i) above). Multi-stakeholder involvement and agreement on processes and methods for RWE use was viewed as important. Similarly, the need for multi decision-maker integration of RWE evaluation was highlighted, including, other than medicines regulators, further engagement with HTA bodies and payers, patients, HCPs, etc.

Other stakeholders could also support planning for artificial intelligence developments and future use with research and methods, as well as other less developed use cases for RWE use and their methodology. Lastly, the group viewed the qualification process for registries as a possible optimisation that registry holders could achieve, to improve the utility of their data and its quality.

iv) What else do you expect from process optimisation, especially in the field of use of RWE?

The session reiterated the need for clarity and guidance – via Q&As, FAQs, best practices – to inform stakeholder what the expectations from regulators are and to reduce the need for (repeat) questions.

Timeliness for regulatory feedback was noted as important. The group also suggested that simplifications of existing processes (procedures) could be considered where legally possible, with e.g. smaller briefing package requirements, in order for advice to be delivered quicker, in a 'less formal', iterative advice (e.g. via SAWP).

The group also expressed the desire for clarity on how and in what structure the provided RWE would be later reflected in EPARs. The group concluded with a strong wish and commitment for continued engagement and follow-up discussion.

4. Breakout session: Training and expertise

Chair: Marilena Vrana, European Heart Network, Brussels; Rapporteur: Valentijn de Jong, EMA

After introductory presentations by Gianmario Candore and Stefania Simou, EMA, the group addressed 4 questions:

i) *Which learning and skills gaps should be addressed in priority to develop the capability of different stakeholder groups to use real-world evidence for regulatory purpose*, e.g. the EU regulatory network, pharmaceutical companies, patients, health care professionals, academic institutions, other stakeholders?

- "People don't know what they don't know"; people need to know which methodological skills are necessary for analysing Big Data but do not know which skills are missing.
- A common understanding of terminology and definitions is missing; the regulatory network needs to be provided or agree on definitions and methods for interpretation of the data and the evidence.
- EMA is using a step-by-step approach and currently uses RWE mainly for drug utilisation, incidence, prevalence and pharmacovigilance studies and less frequently for causal studies. Different methods may be needed depending on the phase in the life-cycle of a medicine.
- Stakeholders welcomed the step-by-step approach and agreed there is a need to investigate which of the methodologies are appropriate and acceptable for the regulators.
- The audience asks regulators to develop a portfolio of methodologies that would be acceptable in the different phases of the drug development and post-marketing authorisation.

ii) *Do stakeholders need training from regulators following publication of guidelines published by EMA and the regulatory network?* Which types of guidelines would require additional training and what type of training material, for example educational material, training sessions, communications, should be used?

- Stakeholders considered they need training on new guidelines, but less so on existing ones, for which they already have more experience.
- "Learning by doing" is key. Sharing real cases and training methodology would be useful to communicate what worked and what didn't work. For instance, the EMA could hold stakeholder meetings for Big Data, similar to those of the EMA Innovation Task Force.

iii) Are you producing training material for your own audience? How could collaborations between stakeholders' groups and academic institutions be best established to fulfil training needs? How could knowledge transfer be organised? How could such interactions be supported by the EU regulatory network?

- Many participants are exposed to a large amount of internal training that is not openly available or can be shared.
- Multi-stakeholder training curricula would be helpful. Regulatory networks and stakeholders could collaborate on training in order to come to a common understanding.
- Keeping the dialog open between regulators and stakeholders is important .
- Accessibility of training materials being developed is crucial.

iv) Could the training curricula on Data sciences, Pharmacoepidemiology and Biostatistics being developed for the EU regulatory network also address the needs of other stakeholders, and through which mechanisms?

- A multi-stakeholder cooperation for the development of such training could consist in multi-stakeholder meetings on real world examples.
- The training materials being developed are currently not widely accessible. An accessible repository of training materials focussing on real world examples would be helpful.

5. Breakout session: Heterogeneity of results between data sources

Chair: Daniel Prieto Alhambra, Oxford University and Erasmus Medical Centre, Rotterdam; Rapporteur: Alexandra Pacurariu, EMA

The objective of the breakout session was to determine how should heterogeneity between RWD sources be analysed and interpreted for regulatory submission and decision-making. Heterogeneity is not a new topic and several review papers addressed this issue already (e.g.Madigan et al.[3]).

After an introductory presentation by Catherine Cohet, EMA, four questions were addressed.

i) Is heterogeneity between results of database studies a common feature in pharmacoepidemiology? What are possible explanations for such heterogeneity? Can heterogeneity be a source of knowledge?

- Heterogeneity is a common problem when analysing multiple databases. In the past, the risk of heterogeneity has been interpreted as a lack of reliability or bias and this was used as an argument against real world data. Heterogeneity should however be appreciated as a source of knowledge and understanding of the complexity of different settings for generating RWE, and therefore it should not be necessarily removed or corrected for.
- True heterogeneity should be distinguished from artefactual heterogeneity. Four types of heterogeneity exist: true heterogeneity, measurement heterogeneity, information heterogeneity (e.g. due to the availability and granularity of information) and interpretation heterogeneity (e.g. statistical significance as a metric for trial emulation). Any source of measurement, information or interpretation heterogeneity due to design errors needs to be eliminated. It can stem from how data is collected and reported and sometimes due to constraints of collection. Conversely, *true heterogeneity* is potentially valuable and should be preserved.

[3] Madigan D. et al. Evaluating the impact of database heterogeneity on observational study results. Am J Epidemiol. 2013;178(4):645-51

- Different types of data may be needed to learn on heterogeneity and sets of metrics are needed to measure database heterogeneity and consistency of the results**.**

ii) Can heterogeneity between data sources be anticipated when considering use of a range of databases, e.g. through set of standard indicators (metadata) or other means?

- Metadata can be useful to anticipate heterogeneity but cannot solve the issue completely.
- Phenotype libraries have a potential value to identify important variables in different databases.
- Cohort diagnostics, analytical diagnostics and summary scores (e.g. propensity scores) are valuable tools that could be implemented before conducting a study.
- Data quality assessments and periodicity should be documented.

iii) What are possible remedies to attenuate heterogeneity at the stage of study design, for example through restriction of study population, exposure and outcomes to the minimum required?

- Restriction of the study population is appropriate for certain questions but has a very high cost for RWE as it potentially sacrifices statistical power and external validity.
- The group considered that more useful solutions could include the use of metadata to identify 'fit for purpose' databases for a specific research question and the use of negative control outcomes to address systematic error, residual confounding and other sources of bias.
- Appropriate choices of study design and visualisation tools are key to minimise "unwanted" heterogeneity.

iv) How can heterogeneity between databases be analysed and interpreted in the context of regulatory evaluations? Is there a place for (meta-)analytical techniques? Under which conditions?

- The answer depends on the source of heterogeneity. Pooling of individual patient data would be desirable but is difficult to apply in Europe due to data privacy issues. Having access to patient-level data for such studies and analyses should be considered of public health importance.
- Visualisation tools can be used to identify differences in study design and to interpret a potential source of heterogeneity. Forest plots before meta-analysis are very informative.
- All estimates should be reported by using forest plots; results should not consist only in the estimates of meta-analyses as a meta-analysis can discard useful data by requesting them to be homogenous.
- The interpretation of results should take into account that heterogeneous findings may be true and reliable.
- The specific issue of small cell counts should be carefully considered. While zero outcomes have a meaning, as they may mean "safe" in a safety study, such data are usually dropped in meta-analyses for mathematical reasons. The Maentel-Hanszel approach and some novel analytical/IT solutions e.g. use of synthetic data, have been used and could be explored as possible solutions.

Plenary session 3: Perspectives and next steps

Chair: Jesper Kjaer, Danish Medicines Agency

1. Patrick Ryan, from Jansen Research and Development and Columbia University Irving Medical Centre, the United-States, provided a **5-year outlook for the analysis and interpretation of real-world data**. Starting from the EMA's vision that "*by 2025 the use of RWE will have been*

enabled and its value will have been established across the spectrum of regulatory use cases" [4] Patrick Ryan highlighted that ensuring the safe and effective use of medicinal products and the appropriate use of RWE to inform regulatory decision-making is not just a European regulatory responsibility but is a global responsibility for all stakeholders. A challenge is however the current status-quo in observational research based on the traditional sequence used for conducting studies: *curate data*, *select cohorts*, *implement analysis* and *disseminate evidence*. The reliability of the evidence is often questioned based on possible concerns that are both methodological (data quality, measurement errors, method bias) and technological (correctness of the ETL process, of the logic leading to cohort selection and of the programming of the analysis). Several questions therefore arise: does the study provide an unbiased effect estimate? Are the findings generalisable to the population of interest? Do the results show a consistent effect across the network and how does heterogeneity across the network impact interpretation? Can the study be fully reproduced across the network?

Using a table of desirable attributes for reliable evidence (repeatable, reproducible, replicable, generalisable, robust, calibrated), Patrick Ryan proposed that a system for RWE generation based on a standardised data network can be empirically demonstrated to be reliable, and application of a common data model can enable standardised analytics across a distributed network. The characteristics for such system was presented and included three main steps: *data quality evaluation* (using database diagnostics), *phenotype development and evaluation* (using cohort diagnostics) and *analysis reliability evaluation* (using study diagnostics). Final unblinded results would be displayed in an interface for exploration.

Three complementary types of evidence to generate real-world data were also presented: c*linical characterisation* (observation), *patient-level prediction* (inference) and *population-level effect estimation* (causal inference) These evidence types may be mapped to different regulatory use cases, categorised into "support the planning & validity of applicant studies", "understand clinical context" and "investigate associations and impact." and they may be informed by different research questions.

The level of proactivity in delivering real-world evidence that could be achieved over the next 5 years was discussed. Proactivity was categorised into 5 categories from "Reactive bespoke" (including problem statement, protocol development, statistical programming, results review, report writing and delivery to decision-maker) to "Anticipatory" (incl. generation and delivery of insights without being asked and answering questions before requested by 'pushing' relevant pre-computed evidence to potential evidence consumers). Time-to-evidence would range from "~weeks, months or years" for Reactive Bespoke to "seconds" for Anticipatory levels. Based on published examples, it was estimated that, in 2026, evidence could be generated within the following timelines: hours for "Support the planning & validity of applicant studies" use cases, seconds for "Understand clinical context" use cases and minutes for "Investigate associations and impact" use cases.

Patrick Ryan concluded that:

- enabling use and establishing value of real-world evidence is a reasonable vision which requires building trust across evidence generators and consumers
- people and processes need to be augmented with science, technology and engineering

4 Arlett P et al. Real-World Evidence in EU Medicines Regulation: Enabling Use and Establishing Value. Clin Pharmacol Ther. 2022;111(1):21-23

- community efforts today can enable a more proactive future tomorrow
- open science systems that promote transparency and reproducibility can increase reliability and efficiency
- regulatory use cases largely involve characterisation analyses, have been demonstrated to be feasible and are ready-to-scale.

2. Jesper Kjaer (Danish Medicines Agency) reminded the audience about the Big Data Steering Group (BDSG) Workplan 2021-2023, stressed key progress performed towards the end of 2021 and provided highlights for early 2022 on several topics, incl. DARWIN EU, Data quality & representativeness, Data discoverability, Network capability to analyse, EU Network processes, International initiatives and the EU BD stakeholder implementation forum. He also presented the objectives for the Clusters of Excellence (CoE) which are included in the BDSG workplan: to bring together learnings from national approaches to data analytics to provide recommendations/good practice to the EU regulatory network. These recommendations will be discussed in a paper on the key building blocks to establish analytics CoE: Data access, Legal aspects, Capabilities, Infrastructure, Method development and AI.

3. Peter Arlett (EMA) thanked the presenters and the audience for their active participation and closed the meeting.