June 2025 edition

IQVIA Real World Data Assets

Navigate a crucial, complex ecosystem to generate insights

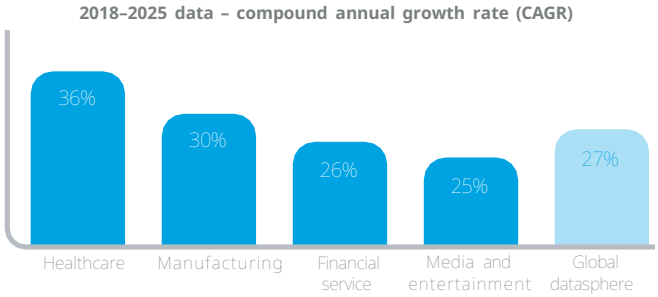
Table of contents

Part One
Maximizing the potential of Real World Data for insights and evidence generation

Traditional Randomized Clinical Trials (RCTs) are widely recognized as a reliable framework for generating evidence on the effectiveness and safety of health technologies for regulatory purposes. However, when alternative sources of evidence, such as Real World Data (RWD) from diverse settings are the only option, the rigid processes of RCTs pose a challenge. Consequently, health authorities and national health systems worldwide are increasingly accepting the need to move beyond relying solely on RCT data.

In the rapidly growing Big Data universe, unprecedented amounts of person-level information contained in Electronic Health Record (EHR) systems contributes to the fastest compound annual growth rate of 36% forecasted till 2025 (growth of volume of health related data).[1] The data originates from primary, specialist and hospital care, drug and disease registries, medical devices, digital apps, and other sources of health data which are generated every day.

VOLUME IS GROWING IN ABSOLUTE TERMS …

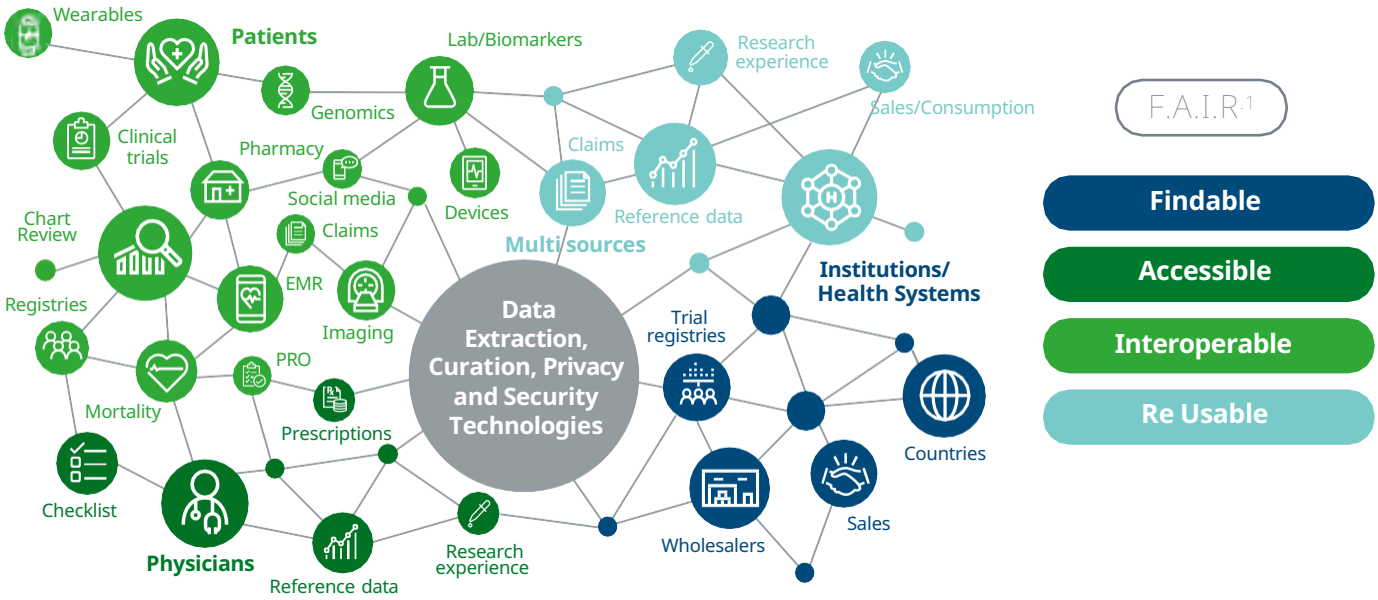


Source: [1] Coughlin et al Internal Medicine Journal article "Looking to tomorrow's healthcare today: a participatory health perspective". IDC White Paper, Doc# US44413318, November 2018: The Digitization of the World – From Edge to Core".

They include structured data in the form of diagnoses, medication, laboratory test results, etc. with different levels of granularity. They also include unstructured data in clinical narratives, for example all of which likely contain invaluable insights for most research projects along the development lifecycle, some of them becoming essential for regulatory purposes. These datasets are often siloed by country, language, region, hospital and even department. Real World data also originate from a multitude of disease specific contexts and so vary greatly in complexity.

As a consequence of this great diversity, it is broadly acknowledged that state-of-the-art tools are needed to make such healthcare data Findable, Accessible, Interoperable and Reusable (as per the FAIR principles initially published in March 2016 by a group of scientists (Mark D. Wilkinson) in Scientific Data); opening the door to maximizing the use of Real-World Data (RWD) in the regulatory context and beyond.

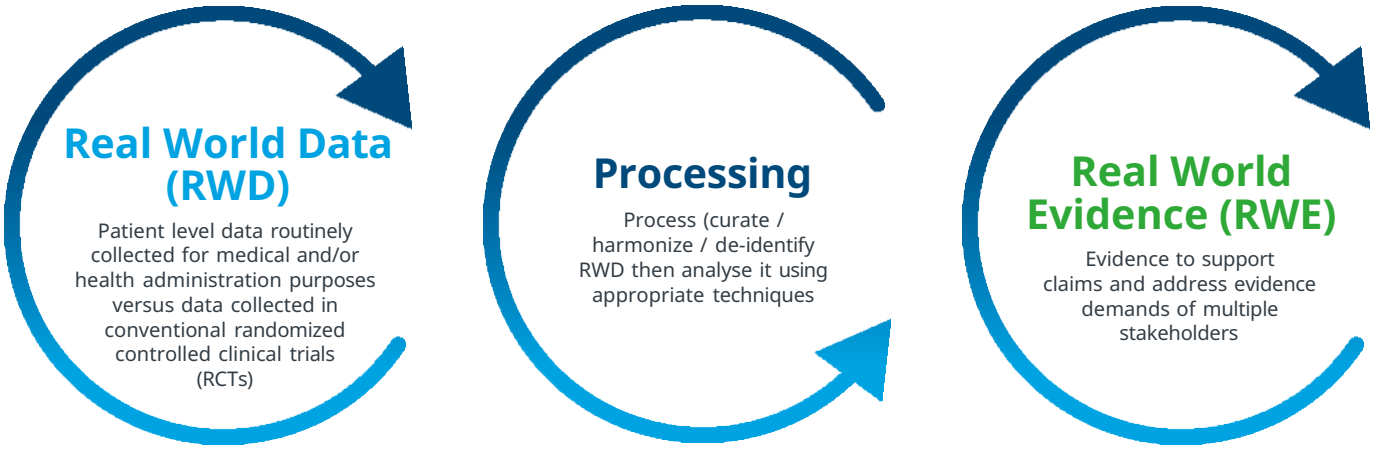
… ALONG WITH CHANGES IN HOW IT IS CREATED



[1] Scientific Data: 15 March 2016 The FAIR Guiding Principles for scientific data management and stewardship – Mark D. Wilkinson

Real World Data and its uses — from Real World Data to Real World Evidence

Real-world data and analytics are generally robust and widely available today across many stakeholders, but generating meaningful evidence remains complex and elusive. This is due to contingent factors, such as the heterogeneous level of data sources quality and the increased usage of AI techniques that are not yet framed by a code of conduct.



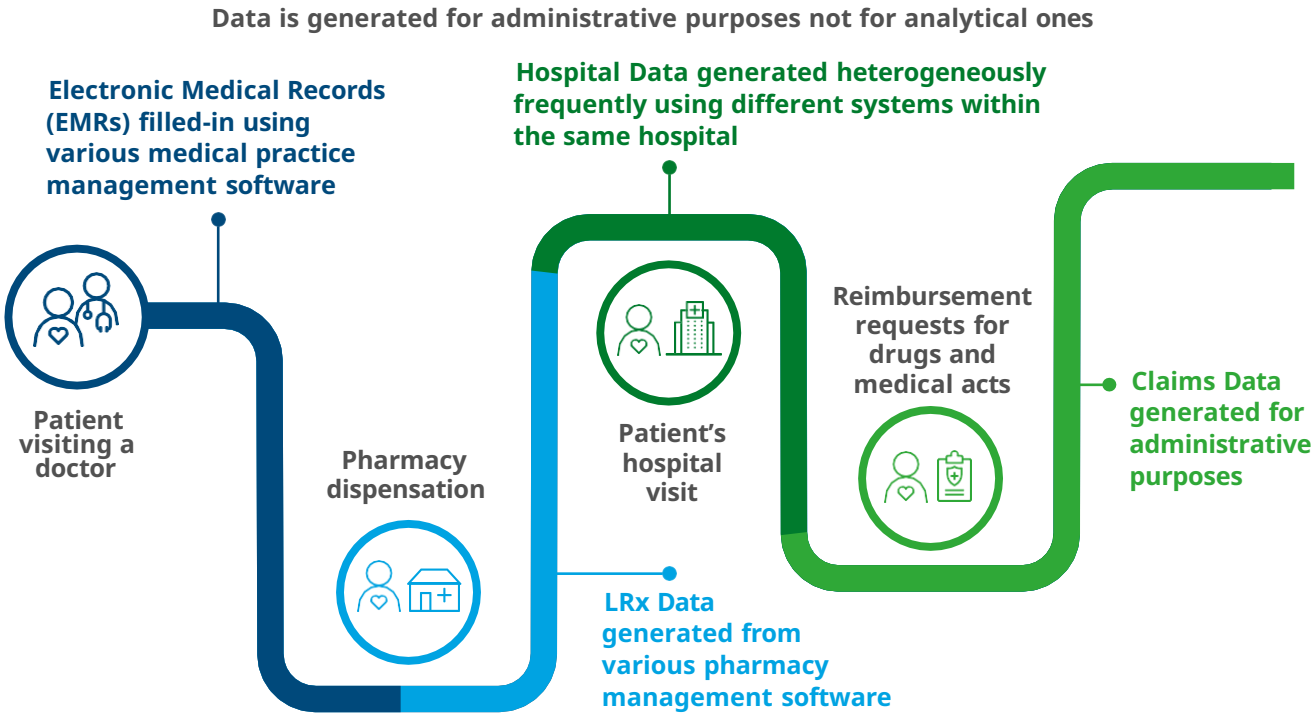
The image above clearly differentiates between Real World Data and Real World Evidence: RWD needs to undergo two types of processes in order to potentially become evidence:

- They initially need to be cleaned, curated, harmonized and de-identified
- They can then be analysed via diverse approaches, ranging from descriptive statistical models to complex Artificial Intelligence algorithms

Focus on main RWD sources

One of the main reasons behind the initial data processing needs outlined above is that the initial intent underlying patient data collection is not to perform analytics (descriptive or predictive) in order to evaluate and support improvement of healthcare management, processes and systems but rather to manage:

- Administrative data for claims, for patients and / or Health Care Professionals to be reimbursed) and pharmacy dispensation (for pharmacists to manage their business and the corresponding supply chain), or
- Patient health records for (EMRs/EHRs) for medical practitioners to appropriately manage their patients' health records



Once curated, harmonized and analytically ready, answer-rich RWD can — and should — be leveraged for multiple usage across the drug development lifecycle. This should be done as early as possible to plan for real world evidence strategy: from adopting the most appropriate clinical trial designs, comparators and evaluation criteria, to performing most thorough post marketing evaluation.

Pharma Companies and regulatory bodies are therefore investing to build internal knowledge and expertise and are adopting RWE- based strategies to support effective and safe product launches, new development.needs and drug use assessment and impact in real world settings.

The next couple of pages provide a topline description of the characteristics, strengths and weaknesses of the above mentioned types of RWD.

Profile of the main Real World Data sources

EMR – ELECTRONIC MEDICAL RECORDS[1]

Office-based HCPs/Outpatients

Available in 10 countries

- GPs and specialists in Canada, France, Germany, Spain, USA
- GPs only in Australia, Belgium, Italy, New Zealand and UK

Notes:
*In **Spain**, EMR data also covers **outpatient** information*

Key information collected

- **Diagnosis**
  - Diagnosis history
- **Patient profile**
  - Demographics: Age, gender
  - Biometrics: Height, weight, BMI
  - Medical history
- **Prescribed treatment**
  - Drug code
  - Script duration, daily dosage
- **Lab results**

Especially appropriate for:

- **Chronic conditions** managed in primary care
  - Diabetes, respiratory, cardiology, neurology, gastroenterology
- **Acute conditions** managed in primary care
  - Flu, bronchitis, headache, gastroenteritis...

Usually not appropriate for

- **Rare diseases** (as often limited national coverage)
- **Hospital-managed conditions**

WHAT	Anonymized patient records collected from Patient Management software used by GPs and selected specialists during an office visit to document patients' clinical records
WHERE	IQVIA collects prescription, patient and medical information from office-based HCPs
HOW	Data from physicians' practices are collected directly from the software, then processed, quality-checked and consolidated
WHEN	Data are collected daily, consolidated monthly
FOR	Brand performance, market dynamics, patient characteristics, treatment patterns, prescribing behavior, prescriber profile Pharmaco-epidemiology, PASS/DUS studies, PV, HEOR, compliance & persistence, pharmaceutical guidelines, drug usage, ...
NOT FOR	PROs, QoL

LONGITUDINAL PRESCRIPTION (LRX)[1]

Outpatients/General population

- All prescribing specialties, including hospital Rx dispensed in retail pharmacies
- Mostly reimbursed treatments
- **Available in 20 countries** (Australia, Belgium, Bulgaria, Canada, Denmark, Finland, France, Germany, Hungary, Ireland, Japan, Kazakhstan, Netherlands, New Zealand, Poland, South Korea, Sweden, Switzerland, UK, USA)

Notes:
*in **Bulgaria**, **Hungary**, **Kazakhstan**, **Netherlands** (partially) and **Sweden**, LRx data include **diagnosis**.*

Key information collected

- **Patient profile**
  - Demographics: Age, gender
- **Dispensed Treatment**
  - Drug
  - Pack size, strength
  - Script duration, daily dosage (inferred)

Especially appropriate for:

- **All prescribing specialties,** including hospital Rx delivered in the retail setting in most countries

Usually not appropriate for diagnosis information (no diagnosis, though diagnosis could be inferred via treatments)

WHAT	Longitudinal prescription data track what therapy a patient starts on and how it changes over time
WHERE	IQVIA collects computerized dispensed prescription records at the anonymized patient level collected from retail pharmacies or coding centers
HOW	Longitudinal prescriptions are collected directly from the retail pharmacies and coding centers
WHEN	Data are typically reported monthly or quarterly
FOR	Brand performance, market dynamics, patient pathway, treatment patterns, concomitant drug use, compliance and persistence, ...
NOT FOR	Clinical effectiveness, PROs, QoL

[1] Acronyms definitions: BMI (Body Mass Index), DUS (Drug Utilization Study), GP (General Practitioner), HCP (Health Care Professional), HEOR (Health Economics Outcomes Research), PASS (Post Authorization Safety Study), PRO (Patient Reported Outcome), PV (Pharmacovigilance), QoL (Quality of Life), Rx (Medical Prescription).

Illustrated by IQVIA's RWD portfolio of proprietary data sources or close collaborations.

HOSPITAL DATA[1]

- Usually based on **hospital administrative data** collected for the purpose of billing & paying for health services or to monitor health care provided
- **Available in 14 countries** (Belgium, China, Germany, Israel, Japan, Mexico, Netherlands, Poland, Portugal, south Korea, Spain, Taiwan, UK, USA)
- Covering inpatients and hospital outpatients **as well as ICU and ERs in some cases**

Key information collected

- **Diagnosis**
- **Patient demographic profile**
- **Prescribed treatment**
  - Drug code
  - Drug consumption
- **Procedures performed**
- **Hospital stay**
  - Hospital characteristics
  - Length of stay
  - Source of admission
  - Locations of care

Especially appropriate for:

- **All types of hospital-managed conditions,** provided they can be identified with ICD10

Usually not appropriate for

- **Primary care conditions**
- Conditions identified via a very detailed ICD10 (> 4 digits)

WHAT	IQVIA's Hospital data assets provide unique insights into what happens and how patients are being treated during hospital visits
WHERE	Data are collected from hospital administrative data such as MBDS (Minimum Basic Data Set), either directly or through TTPs
HOW	IQVIA hospital data is an aggregation of administrative data provided by individual hospitals for administrative/accounting purpose
WHEN	Update frequency varies across data sets, from monthly to yearly
FOR	Pharmaco-epidemiology, PASS studies, PV, HEOR, compliance & persistence, pharmaceutical guidelines, prescribing behavior, prescriber profile, drug usage, market dynamics, …
NOT FOR	HCRU / costs, clinical effectiveness, PROs, QoL

CLAIMS DATA[1]

- Either **adjudicated claims** (from payers) or **pre-adjudicated/ medical claims** (from HCPs)
- **Available in 17 countries** (Australia, Brazil, Canada, China, Colombia, Czech Republic, France, Germany, Hong Kong, Hungary, Israel, Italy, Japan, Slovakia, South Korea, Taiwan, USA)
- **Usually covering all settings of care**

Key information collected

- **Patient demographic profile**
- **Diagnosis**
- **Dispensed Treatment**
  - Drug
  - Pack size, strength
  - Script duration, daily dosage (inferred)
- **Procedures performed**
- **Cost of services, HCRU**

Especially appropriate for:

- **All types of conditions** (identified with ICD10)

Usually not appropriate for

- Conditions identified via a very detailed ICD10 (> 4 digits)

WHAT	Adjudicated or pre-adjudicated claims collected from payers or HCPs
WHERE	IQVIA claims usually capture data across all settings of care
HOW	The data is captured from health insurers or health care providers
WHEN	Update frequency varies across data sets, from monthly to yearly
FOR	Brand performance, market dynamics, prescribing behavior, prescriber profile Pharmaco-epidemiology, PASS studies,PV, HEOR, compliance & persistence, pharmaceutical guidelines, drug usage, HCRU / costs
NOT FOR	Clinical effectiveness, PROs, QoL

[1] Acronyms definitions: ER (Emergency Room), HCP (Health Care Professional), HCRU (Health Care Resource Utilization), HEOR (Health Economics Outcomes Research), ICD (International Classification of Disease), ICU (Intensive Care Unit) PASS (Post Authorization Safety Study), PRO (Patient Reported Outcome), PV (Pharmacovigilance), QoL (Quality of Life), TTP (Third Trusted Party).

MULTI-OMICS DATA[1]

- Linked clinical-genomic datasets from healthcare systems, biobanks or registries. Genomic profiling may differ e.g. SNP, WES, WGS. Biomarker data is also available in some EMR datasets
- Bio samples are often available for testing / multi-omic sequencing (or prospective sampling can be arranged)
- Linked clinical data is usually SOC but may also include HCRU, imaging & pathology reports
- **IQVIA network provides access to genomic data from 30+ data partners and 4 IQVIA data sources in more than 50 countries**

Clinical variables collected vary heavily but generally:

- **Diagnosis**
- **Patient demographic profile**
- **Prescribed treatment**
  - Drug code
  - Drug consumption
- **Procedures performed**
- **Hospital stay**
  - Hospital characteristics
  - Length of stay
  - Source of admission
  - Locations of care

Especially appropriate for:

- **All types of hospital-managed conditions,** provided they can be identified with ICD10

Usually not appropriate for

- **Primary care conditions**
- Conditions identified via a very detailed ICD10 (> 4 digits)

WHAT	IQVIA has global data partnerships covering all TA's including: oncology, rare disease, CNS, immunology, auto-immune, metabolic, as well as global sequencing & lab capabilities
WHERE	From healthcare systems, biobanks, national programmes, registries and PAGs
HOW	Depending on research questions of interest, IQVIA can advise on the best approach and datasets to consider. IQVIA can conduct research on behalf of clients using blended research teams or facilitate access for research where permissions allow
WHEN	Dependent on data set-clinical data sets usually refreshed. Omics data usually point in time. Imaging may be available
FOR	Molecule to market use cases: translational medicine, clinical trial design, patient recruitment, Pharmaco-epidemiology, genomic registries, sub-cohort profiling, regulatory evidence, payer evidence, unmet need, natural history of disease, patient journey
NOT FOR	Use cases subject to ethical approvals and IRB – usually require protocolised research

[1] Acronyms definitions: ER (Emergency Room), HCP (Health Care Professional), HCRU (Health Care Resource Utilization), HEOR (Health Economics Outcomes Research), ICD (International Classification of Disease), ICU (Intensive Care Unit) PASS (Post Authorization Safety Study), PRO (Patient Reported Outcome), PV (Pharmacovigilance), QoL (Quality of Life), TTP (Third Trusted Party).

Real World Data possible uses

As mentioned above, RWD represents a heavy investment to become analytically ready and potentially generate insights. It should therefore be leveraged as much as possible along the drug development lifecycle.

The image below shows examples of use of RWD to answer a diverse set of business/research questions:

Clinical Trial Optimization

Identify best sites and most appropriate patients. Support protocol design

- Protocol Design & Feasibility
- On-going protocol adjustment
- Leverage RWD for study comparative arm
- Optimise country allocation
- Site & patient selection

Epidemiology Assessment

Monitoring of pathology evolution and therapeutic strategies

- Understand the natural history of disease
- Characterize patient populations & identify subgroups of interest
- Treatment pathway
- Determine the standard of care
- Identify unmet needs
- Identify suitable local comparators
- Patient flow analysis/ patient journey
- Adherence studies
- Off-label use

Drug Safety & Risk Management

Segment, analyze and assess the safety and risk/benefit of therapeutic interventions in a real-world setting

- Signal detection and assessment
- Safety Surveillance
- Vigilance
- Risk Assessment
- DUS (Drug Utilization Study)
- PASS (Post Authorization Safety Study)

HEOR/ Market Access

Demonstrate the value of medicine through evidence-based health economic evaluation and real-world outcomes for optimal pricing, reimbursement and coverage potential

- Cost of Illness/HCRU (Health Care Resource Utilization)
- Burden of Disease
- Budget Impact
- Outcomes studies
- Comparative Effectiveness
- Compliance & Persistence
- Contract Optimization
- Target population

Commercial Analytics

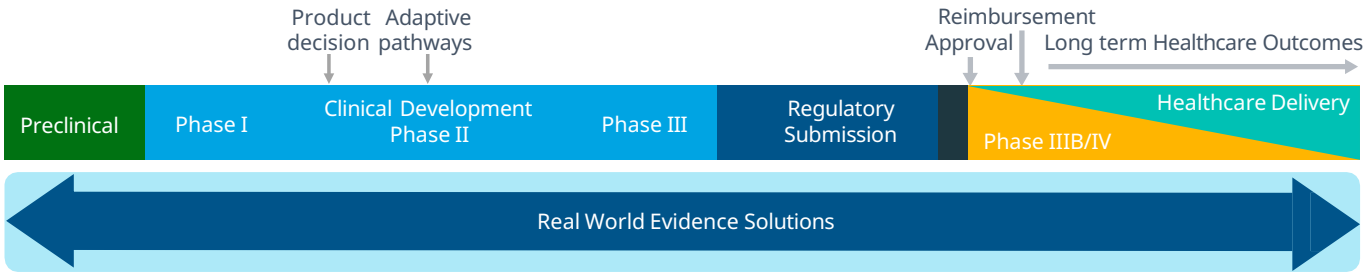
Diagnose, plan,forecast and track brand performance. Size and characterize the target market from the disease and treatment pattern perspective

- Brand/launch Planning & Strategy
- Market sizing and forecasting
- Brand Diagnostics
- Brand Performance tracking/Source of business
- Split by indication
- Contract Compliance

Over the past decade, healthcare stakeholders have come to appreciate the value proposition of RWE benefits. It is increasingly recognized as a powerful tool to generate uniquely valuable evidence thereby supporting healthcare product development and lifecycle strategy.

Regulatory bodies still need to further adapt their processes supporting market approval to encompass RWD generated insights, thus complementing the gold standard of prospective randomized controlled clinical trials (RCTs). It is now recognized that, treatments for certain indications such as rare diseases can only be evaluated on the basis of RWD.

EXAMPLES OF RWE THROUGHOUT THE PRODUCT LIFECYCLE



Increasing role of RWD to potentially generate insights or evidence through-out the product lifecycle

Examples uses in early development

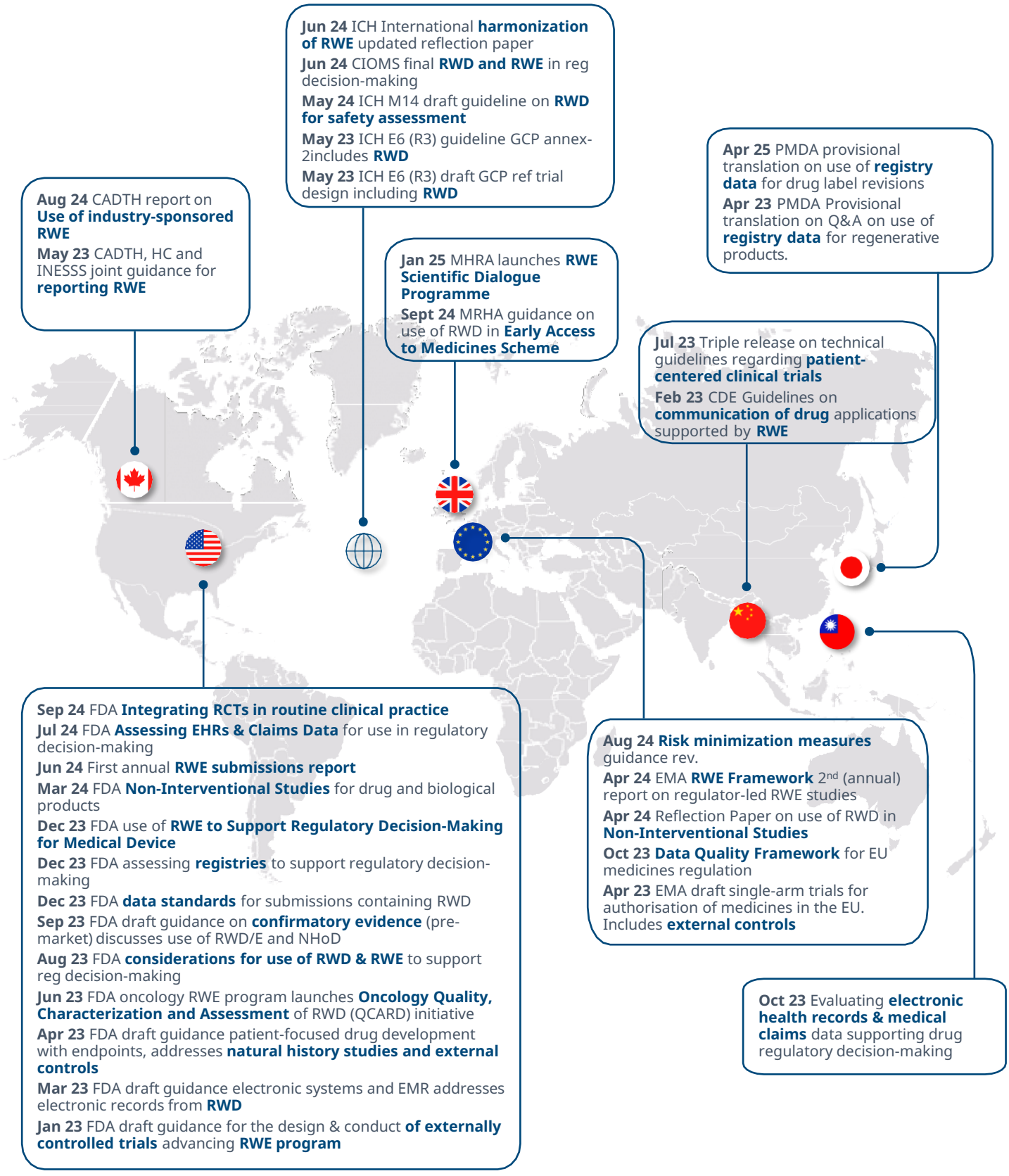
- Evidence planning; Data strategy plan
- External Comparators
- Validate biomarkers and other potential outcomes
- Disease Registries
- Regulatory Intelligence to inform indication and dosage strategy

RWD for RWE post approval

- Demonstrate value & differentiation
- Labeling expansion
- Real-world safety & effectiveness
- Patient perspective/PRO
- Cost-effectiveness
- Implications for Rare Disease (Registry; SOC, Feed Recruitment of Ph2/3

In this context - even though real world data is not generated within an as reassuringly structured and monitored process -, regulators across the globe increasingly consider Real World Evidence in their decisions.

Below some examples of regulations or guidances recently published (Highlights Jan. 2023 – Sept. 2024*)



*Note: Selected citations, not exhaustive

= International application

RWD in a nutshell

EXAMPLES OF USE	DATA TYPES	ACCESS MODELS
Selected examples of use: • **Clinical trial optimization:** inform optimal sites and patients identification. Support protocol design • **Epidemiology assessment:** monitor pathology evolution and therapeutic strategies • **Drug Safety & Risk Management:** assess safety and risk/benefit of therapeutic interventions in the real-world. **Inform Covid-19 response** • **HEOR/Market Access:** demonstrate the value of medicine through evidence based health economic evaluation and real-world outcomes • **Commercial Analytics:** analyze, plan and forecast brand and market	**Main types of patient databases available:** • **Longitudinal prescriptions (LRx):** from pharmacy channel • **Electronical Medical Records (EMR):** from medical practice management software • **Claims:** from Payer, Medical, Hospital, Drug Plans • **Hospital Patient Encounters:** in-patient treatment, outpatient (in some countries) • **Registries and patients' cohorts:** typically generated by academic research • **Cross sectional data:** patients' charts (when longitudinal data does not exist) • **Genomics data:** global access to a network of genomic-clinical data	**Main types of access available:** • **Direct:** clients can access data, either at transactional level (DT – Direct Transactional) data displayed on a non-identified patient basis or at query level (DQ – Direct Query) data displayed as aggregate only. Use restrictions may apply • **Indirect (or project per project):** client interactive analytics via software hosted by IQVIA, in a compliant infrastructure • **Federated (or remote):** rapid analytics via standard methods on remote data (e.g. OMOP)

Data type relevance by research question

The table below shows which research questions can be answered by each data type, highlighting strengths and limitations.

	LRx	EMRs	CLAIMS	HOSPITAL
BRAND PERFORMANCE, MARKET DYNAMICS	🟢	🟢🟠	🟢	🟠
PATIENT CHARACTERISTICS	🟠 Demographics	🟢 Demographics, biometrics, medical, clinical, …	🟠 Demographics	🟠 Demographics
PATIENT PATHWAY	🟢 Capturing all dispensations in retail pharmacies, including hospital Rx delivered out of hospitals	🟢🟠 Relevant specialties may not be covered, limited linkage possibilities across specialties	🟢 Covering all settings of care	🟠 Limited to hospital pathway
TREATMENT PATTERNS	🟠 No direct information, can be inferred (daily dosage, length of treatment) in some cases	🟢 Direct access to daily dosage, length of treatment	🟠 No direct information, daily dosage, length of treatment can be inferred	🟠 No direct information, daily dosage, length of treatment can be inferred
HCRU (Healthcare Resource Utilization)/ Costs	🟠 Limited to drug consumption	🟠 Limited to primary care consultations (GPs and specialists) and drug theoretical consumption	🟢 Covering all settings and types of care	🟠 Limited to hospital care
CLINICAL EFFECTIVENESS	🔴 No clinical information directly accessible	🟢 Varies across datasets, therapeutic areas, tests, …	🔴 No clinical information directly accessible	🔴 No clinical information directly accessible
PROs (Patient Reported Outcomes)	🔴 Patient reported outcomes are a great source of data enrichment when secondary data sources allow for enriched studies to be performed (linkage of PRO data to data source; please see IQVIA eCOA Offering).			
QOL (Quality of Life)	🔴 Quality of life data is often obtained via certified formularies which can greatly enrich secondary data sources when possible (please see IQVIA eCOA Offering).			
SETTING UP COMPARATOR ARMS FOR CLINICAL STUDIES	Depending on inclusion criteria	Depending on inclusion criteria	Depending on inclusion criteria	Depending on inclusion criteria
MAIN ADVANTAGES	Country coverage (20 countries) National coverage (usually > 50%) **Specialty coverage**	Patient profile (demographics, biometrics, medical history, …) Medical and clinical information (diagnoses, lab test results, …)	Settings of care coverage (usually primary and secondary care) Diagnosis info (except in some cases)	Detailed hospital information
MAIN LIMITATIONS	Information limited to drug use No diagnosis, medical or clinical info No intra-hospital coverage	No intra-hospital coverage Usually no linkage across specialties	No or limited medical and clinical information in some cases Update frequency Protocolized access (in some countries)	Update frequency No or limited hospital treatment info Protocolized access (in some countries)

🟢 Relevant 🟠 Partially relevant, potential limitations 🔴 Not relevant

Real World Data challenges

Addressing some of the FAIR principles

As previously illustrated, most of the challenges encountered when transforming Real World Data into Real World Evidence are due to the heterogeneity of data sources in different settings, entered in various formats and with differing access challenges. A Healthcare Professionals' first focus is to treat patients, not to perform data collection, and even less so to perform it in a homogeneous manner.

The FAIR Data Principles (Findable, Accessible, Interoperable, and Reusable), published in Scientific Data in 2016, are a set of guiding principles proposed by a consortium of scientists and organizations to support the reusability of digital assets. They represent a useful framework when envisaging sharing data in a way that will enable maximum use and reuse.

In this section, we are adressing some of the FAIR principles from an operational/functional point of view rather than a technical one, illustrating them with chosen tools/platforms and processes.

The FAIR principles highlight the challenges of:

- Finding the fit for purpose data assets
- Accessing those sources
- Implementing the right/increasing level of interoperability
- Establishing the framework to reuse data generated

Find

Need for:

- Dynamically updated / orchestrated syndicated catalog
- Robust processes to select fit for purpose data

Interoperate

Enable data sources comparability and interchange with clear understanding of applied data transformation and dissemination processes

Access

Facilitate data access accross multiple settings in a transparent and compliant manner

Reuse

After initial collection, reuse data in a compliant/transparent fashion via all statistical/data science to produce better outcomes

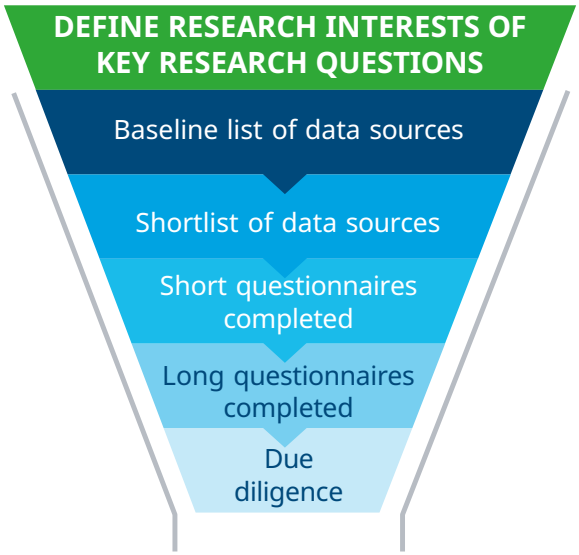
Data findability

A DATA LANDSCAPING PROCESS

Determining which data sources are most likely to be of high value, is not always straightforward. Often, data sources must first be cataloged and evaluated based on past experience, before you can know if they are a fit.

In healthcare, the key factors for approaching such information gathering typically include:

1. An understanding of the research needs for which the data are being evaluated
2. Simple 'rule-out' criteria for excluding data sources which are highly unlikely to be of value
3. Criteria by which to prioritise data sources which are deemed to be potentially of value



Organisations engaged in cataloguing and evaluating data sources typically employ a process based around these components. The process follows a funnel/filter approach as illustrated in this picture.

At each step in the process the number of data sources under consideration typically reduces while the effort applied to the characterisation of each data source increases.

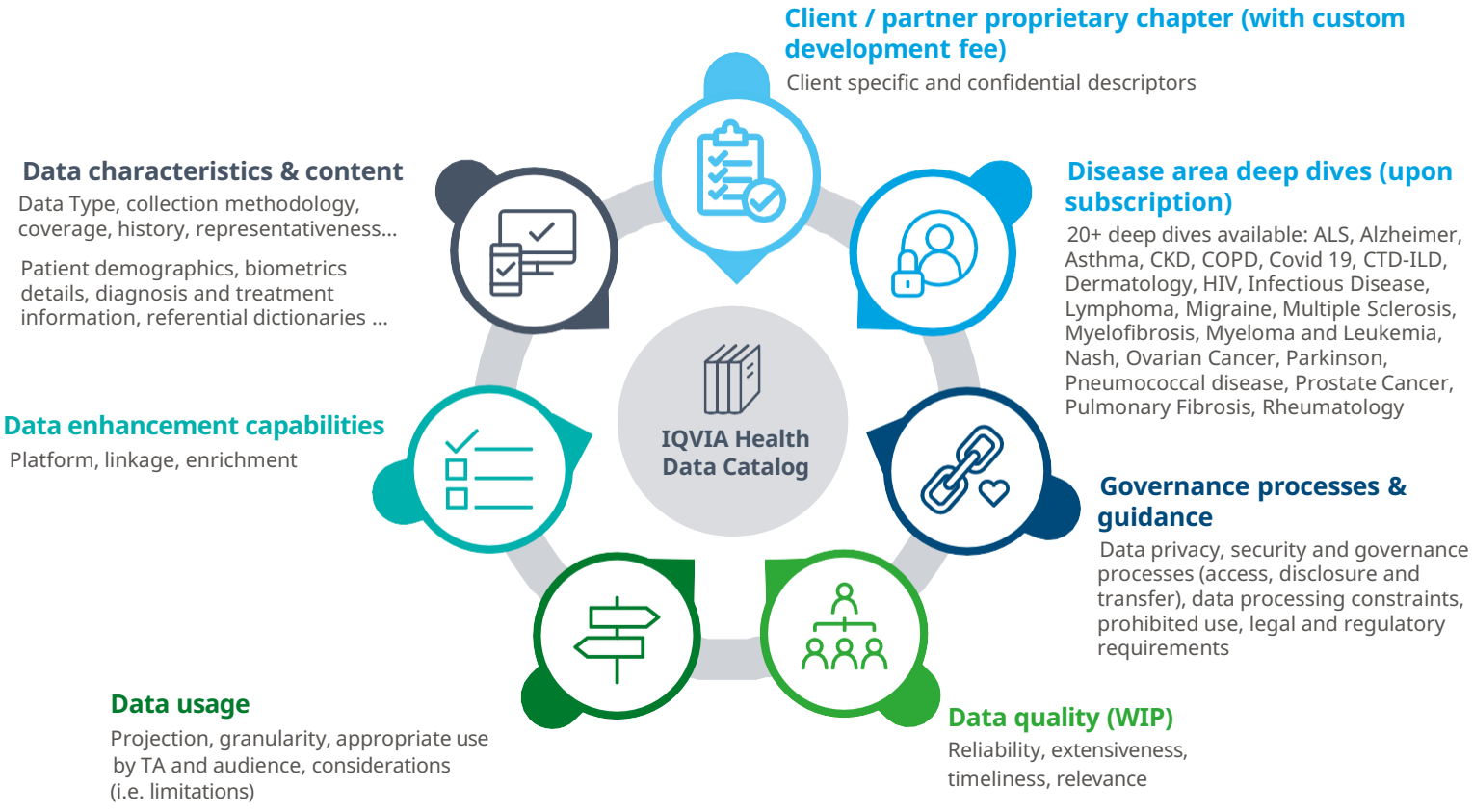
- The early profiling and shortlisting of data sources is often performed by means of a literature search or use of an existing catalogue comprising publicly available information.
- Once data sources are shortlisted for further evaluation, the intent is typically to capture information (metadata) directly from the owners or custodians of those sources by means of structured questions administered either through discussion or as an electronic questionnaire or both.
- The qualitative information captured in this way then typically enables the further prioritisation of data sources based on an assessment of their relative fitness for a particular research need.
- In some cases it may then be appropriate to perform a further quantitative evaluation on the data, deriving measures of completeness and quality. The process follows a funnel/filter approach to select the most appropriate source(s) and inform the design of the research study.

A METADATA CATALOGUE

Describing and evaluating data sources comprehensively is a resource-intensive process. This is especially true if the range of characteristics of interest to be described is broad and the evaluation of specific characteristics is deep. These activities also require the co-operation and engagement of the owners or custodians of data sources who are best placed to provide or validate this information.

IQVIA has developed a Health Data Catalog (IHDC) with unique breadth, depth and governance features to enable first level identification of the most appropriate data sources to be explored to help answer specific research questions.

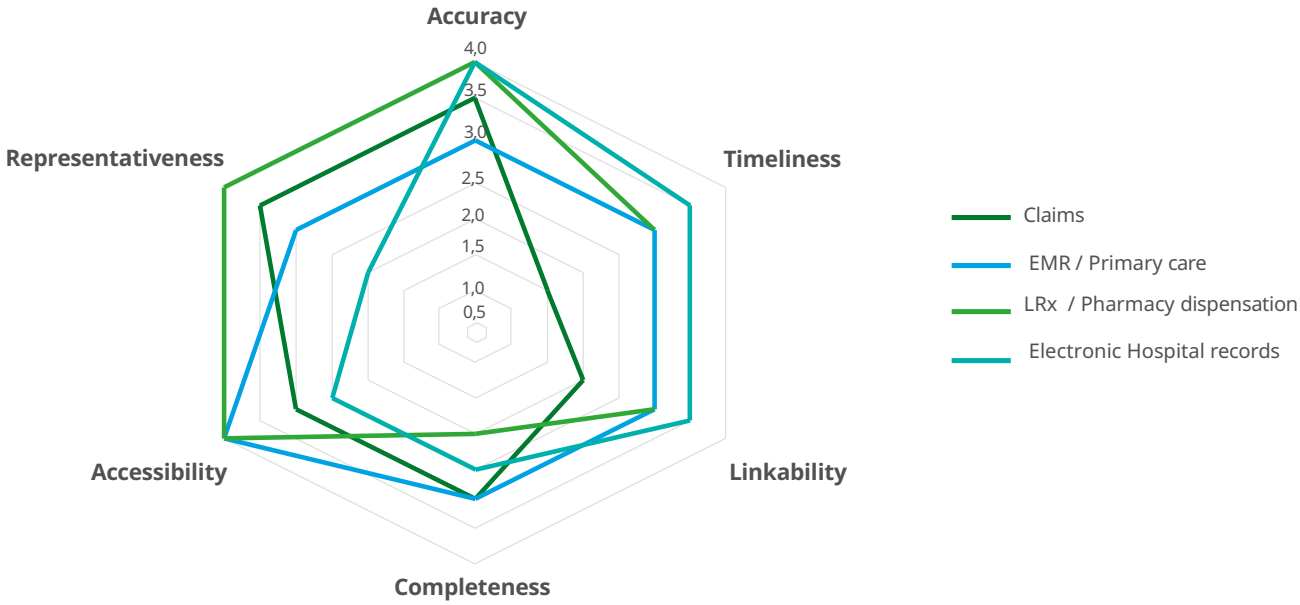
Over 200 variables to describe patient data across 4,300+ data assets



DATA QUALITY FRAMEWORK

Today's healthcare industry relies heavily on the generation and utilisation of data produced in computerised systems, manipulated, aggregated, analysed and visualised for a multitude of purposes. In some cases, these processes produce data, in others, data are an integral part of generation of other products. There are a variety of types of data, depending on their level of organisation, process control and purpose.

The spider net below provides a topline representation of how dataset characteristics can be framed to assess their level of quality and appropriateness for diverse research projects.



CASE STUDY – USE IHDC TO IDENTIFY DBS COVERING LEUKEMIA IN EUROPE AND CAPTURING BIOMARKERS

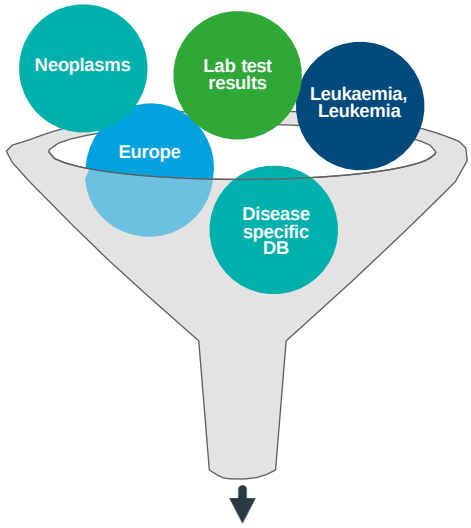
Search by location (Quick search)
Europe

Search by key word (Search box)
Leukaemia, Leukemia

Search by disease area
Search by disease area* (Quick search)
ICD10 category II "Neoplasms"

+ Filter by disease specific DBs (Advanced search)
To exclude general DBs such as EMRs or Pharmacy dispensing (LRx) data.

Search by variables of interest (Advanced search)
Does contain **lab test results**

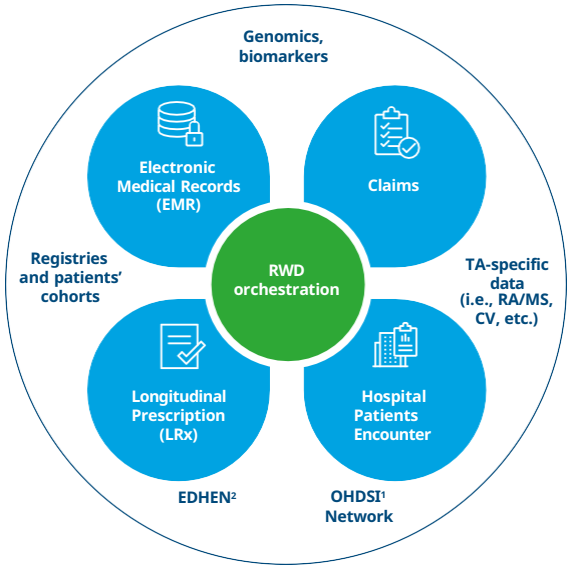


210+ DBs*
registries, research/clinical DBs, patient chart studies, etc.

* Indicative figures, may change over time as the content of the catalog is continuously improving

DATA ORCHESTRATION

While numerous public/private initiatives are underway to render real world data easier to **Find, Access, Interoperate** (=process across many different settings) and **Reuse,** different RWD sources each come with their strengths and limitations, therefore it may be the case that one RWD source does not fully answer to a specific research question.



- Selecting the right data source is of the essence to ensure researchers have fit-for-purpose data to answer the research question at hand
- Understanding the **differences in origin and collection methods** of the most common sources of RWD is an important first step in this process[1]
- Then identifying **opportunities to combine or potentially link** at patient level, different sources may allow to overcome potential data gaps and provide richer insights

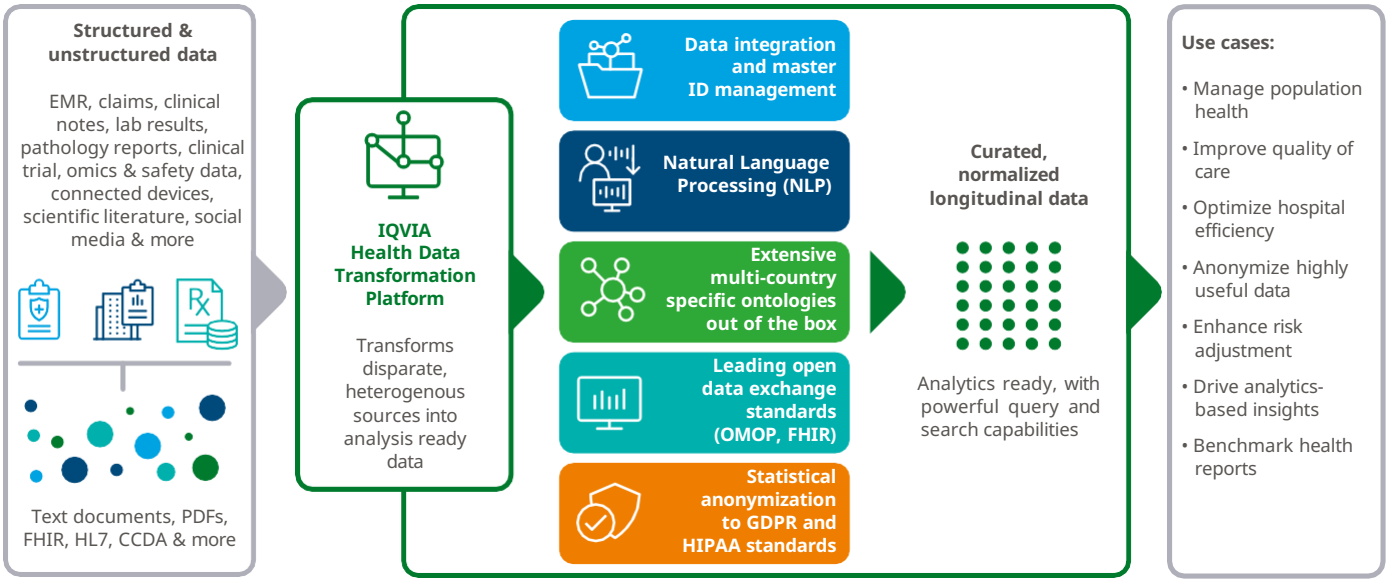
Identifying and managing combination of aims to identify and manage any combination of data/access types in compliance with data protection and privacy rules is probably the biggest challenge when needing to leverage RWD.

[1] OHDSI- Observational Health Data Sciences and Informatics is a public initiative independent from IQVIA, enabling faster, more reliable studies and/or expanded data access via the OHDSI global networks
[2] EHDEN- European Health Data and Evidence Network

Data accessibility and interoperability

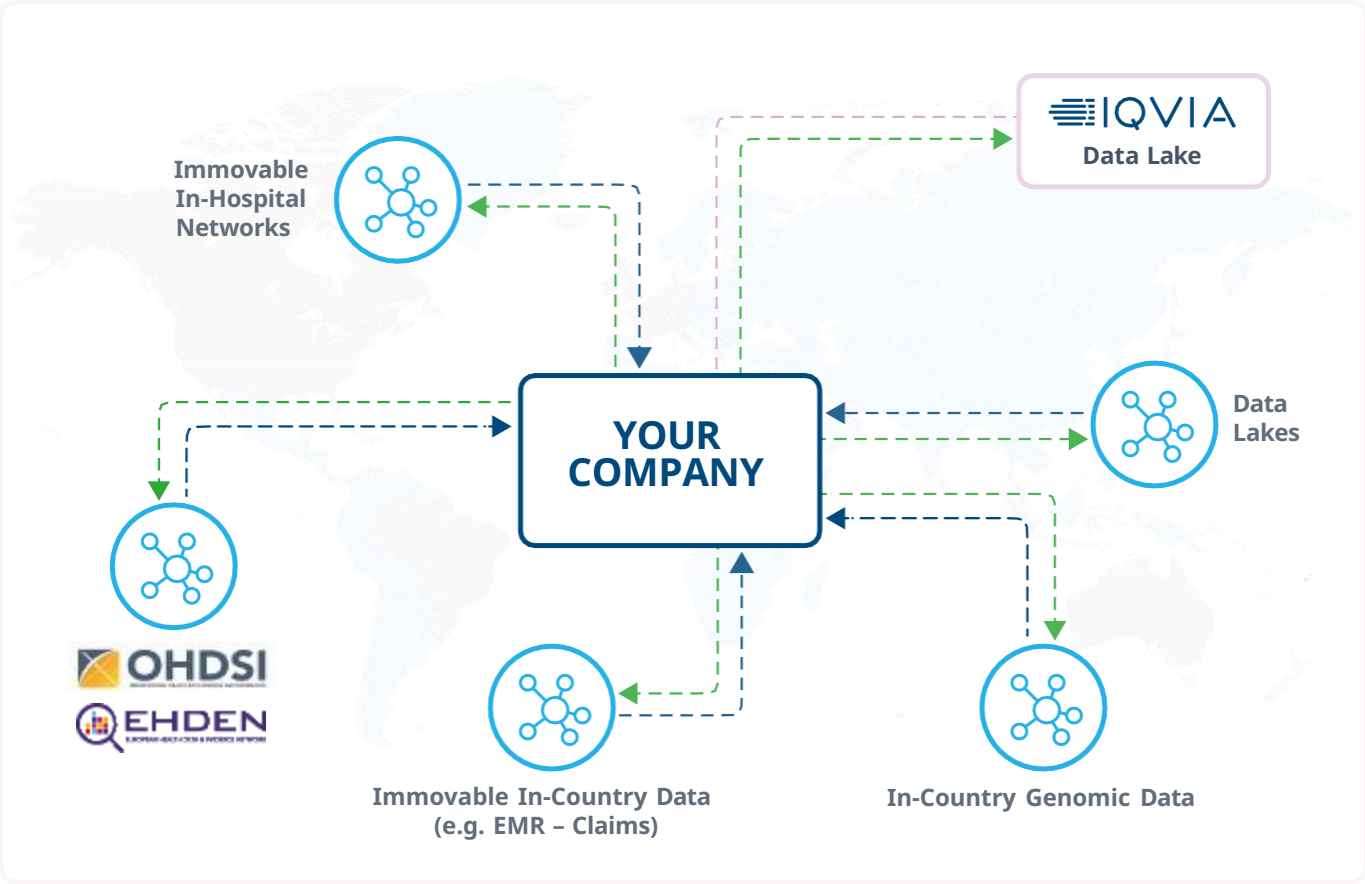
DATA TRANSFORMATION / PROCESSING

Before being accessible, real world data — not collected initially for analytical purposes – needs to undergo transformation processes so that it may become analytically ready. Generating such appropriate Real World Data in view of potentially becoming Real World Evidence is not trivial. It requires secure infrastructure, state of the art data production factories to integrate, clean, curate, bridge, de-identify and perform ongoing quality controls on data collected from diverse sources… As illustrated in the IQVIA example below: heavy investments in infrsatructure and proecesses need to be envisaged.



ACCESSING DATA WHEREEVER IT IS

Current market trends, augmented by privacy concerns, highlight an increased reluctance of providers to let go of the data they process and indicate that RWD "travels" less. Data access is more and more done using a federated model: decentralising analysis to be performed on data sets in a homogeneous manner wherever they are behind firewalls.



Source: The variety of data types and access models makes data orchestration capabilities and expertise key factors of success. According to project's objectives, RWD subject matter experts need to identify best combinations or linkage of data assets or access to partner networks such as OHDSI

Interoperability is the ability of two or more systems to exchange health information and use the information once it is received. It will take time for all types of health IT to be fully interoperable. Meanwhile converting patient data into common data models significantly enhances the possibility to perform studies/analyses on data originating from different settings.

The growing network of OHDSI members' databases consists of more than 300 datasets, across 30+ countries with 2.7B+ patient records; most of which are commercially unavailable data.

OHDSI (OBSERVATIONAL HEALTH DATA SCIENCE AND INFORMATICS): AN ENABLER OF DATA INTEROPERABILITY(OHDSI)* is an international open-science research collaborative that enables higher quality, reproducible studies through standard analytical methods, tools, the **OMOP Common Data Model (CDM)**** and Standardized Vocabularies.

MARKET ACCESS, REAL WORLD EVIDENCE, HEOR

An OMOP case study: Analysis of dual combination therapies used in treatment of hypertension in a multinational cohort

Findings: In this cohort study of 970 335 individuals from 11 large databases, 12 dual combinations of antihypertensive drug classes were commonly used, with large variation across countries and demographic groups.

JAMA Network Open. 2022;5(3): e223877. https://doi:10.1001/jamanetworkopen.2022.3877

AUTHOR(S)[1]
Yuan Lu, ScD[1,2]; Mui Van Zandt, BS[3]; Yun Liu, PhD[4]; et al.

IMPORTANCE
More than 1 billion adults have hypertension globally, of whom 70% cannot achieve their hypertension control goal with monotherapy alone. Data are lacking on clinical use patterns of dual combination therapies prescribed to patients who escalate from monotherapy.

OBJECTIVES
To investigate the most common dual combinations prescribed for treatment escalation in different countries and how treatment use varies by age, sex, and history of cardiovascular disease.

METHODOLOGY
This cohort study used data from 11 electronic health record databases that cover 118 million patients across 8 countries and regions between January 2000 and December 2019.

Included participants were adult patients (ages 18 years) who newly initiated antihypertensive dual combination therapy after escalating from monotherapy. There were 2 databases included for 3 countries: the Iqvia Longitudinal Patient Database (LPD) Australia and Electronic Practice-based Research Network 2019 linked data set from South Western Sydney Local Health District (ePBRN SWSLHD) from Australia, Ajou University School of Medicine (AUSOM) and Kyung Hee University Hospital (KHMC) databases from South Korea, and Khoo Teck Puat Hospital (KTPH) and National University Hospital (NUH) databases from Singapore. Data were analyzed from June 2020 through August 2021.

EXPOSURES
Treatment with dual combinations of the 4 most commonly used antihypertensive drug classes (angiotensin-converting enzyme inhibitor [ACEI] or angiotensin receptor blocker [ARB]; calcium channel blocker [CCB]; β-blocker; and thiazide or thiazide-like diuretic).

CONCLUSIONS
In this study, large variation in the transition between monotherapy and dual combination therapy for hypertension was observed across countries and by demographic group. These findings suggest that future research may be needed to investigate what dual combinations are associated with best outcomes for which patient.

DATABASE SOURCE Data were extracted from 11 sources of Electronic Health Records from 8 countries: Australia, South Korea, Singapore, China, Taiwan, France, Italy, and the USA.

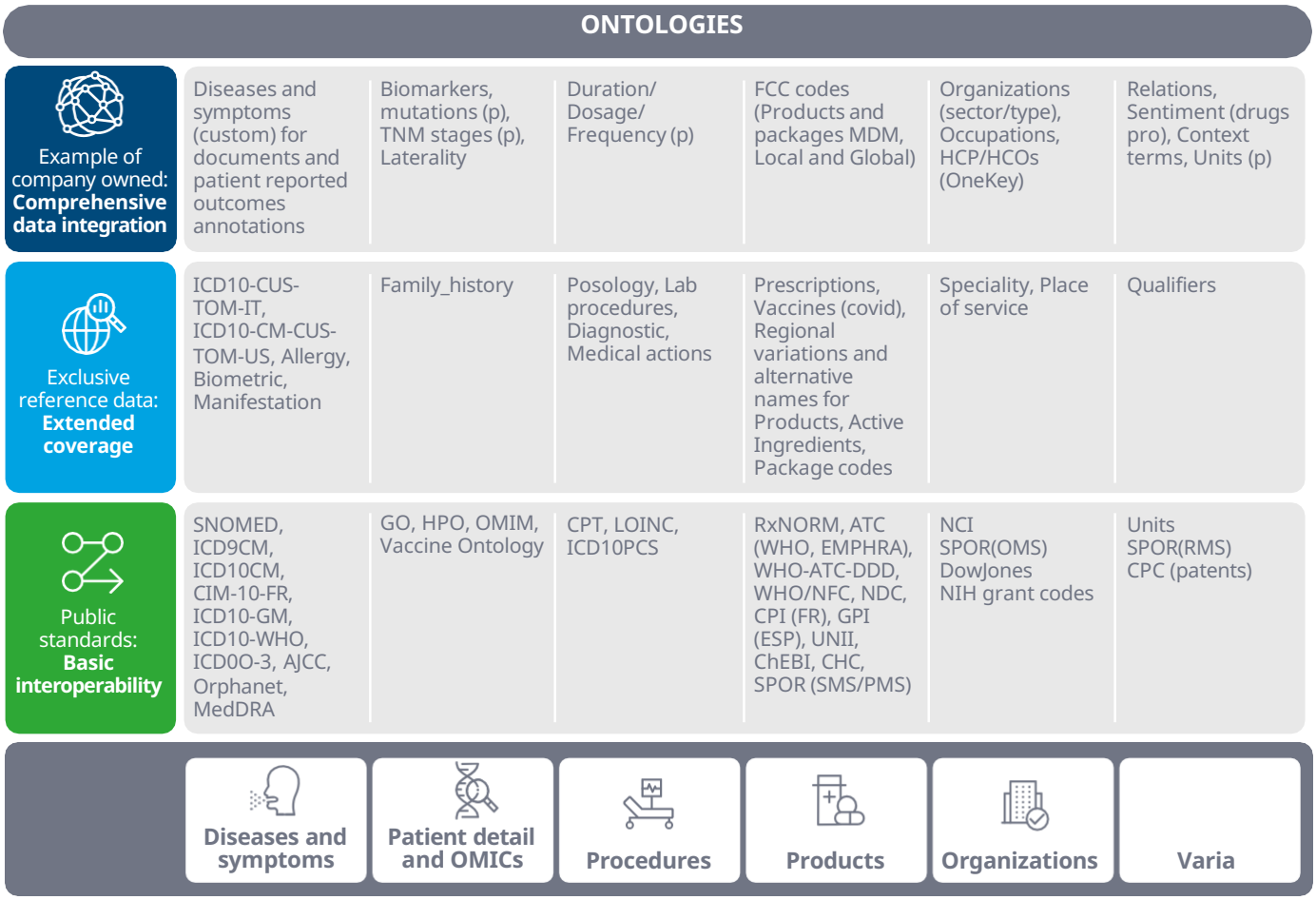
[1] Center for Outcomes Research and Evaluation, Yale New Haven Hospital, New Haven, Connecticut, [2] Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut, [3]Real World Solutions, Iqvia, Durham, North Carolina, [4] School of Biomedical Engineering and Informatics, Department of Medical Informatics, Nanjing Medical University, Jiangsu, China

Data interoperability and reusability

NORMALIZING AND UNIFYING DISPARATE DATA ASSETS FOR GREATER INSIGHTS

Today's complex questions requires staying up-to-date with information across public, proprietary, and subscription-based data. Collecting the information locked in text requires a common language to link terms and meaning regardless of how it's initially written. Many commonly used ontologies as well as proprietary subject specific and linguistic- based ones can be developed and leveraged to convert heterogeneous information into a common language as per the figure below. This enables terminologies with the same meaning to be mapped across data assets.

ILLUSTRATION OF ONTOLOGY ASSETS FOR SELECTED RESOURCES AND DOMAINS

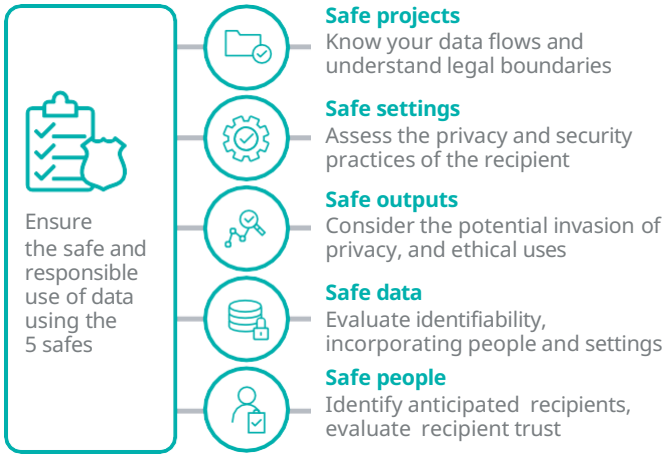


COMPLIANCE AND PRIVACY CHALLENGES

Enable data reuse in compliant settings

From collection to analysis, it is of the essence to perform all data handling/management/processing tasks in a compliant manner in order to enable approriate use and reuse of health data.

The "5 safes"[1] provide a framework of the essential components for the safe and responsible use of data as shown in the picture opposite.



Below we focus on the "Safe people" portion of the 5 safes. "Safe people" is of vital importance to any organization in order to have an empowered workforce who know how to ensure data is suitably protected whilst being used to gain essential insights. People need to be aware of the type of data they are using and what their responsibilities are to protect this data as the use of data comes with the need to manage significant risks. In fact when staff use data inappropriately, it can have serious consequences. These can include fines, removal of access to data, potential of reputational damage and loss of business.

Employees using data need access to organization policies, standard operating procedures and training to educate them to operate in a legal and consistent manner. They also need to be able, and encouraged to reach out to privacy and legal professionals when needed. Typical responsibilities which data users need to be aware of are listed below:

Risk	Responsibility
1. Risks using data	All users need to understand the risks using data, along with their causes and prevention, before gaining access to data including: A. risk of identification B. inference risk C. risk of loss of utility and accuracy D. risk that data is used for illegitimate purposes E. risk of loss of business sensitive data
2. Purposes	Data may only be used for appropriate specified purposes which are agreed in contracts and consent notices. Use of data outside these purposes is not allowed.
3. Use of data	Data use should be appropriate, legal, fair, ethical and transparent.
4. Access to data	Access to data should only be provided on a need-to-know basis for the duration of time needed to complete stated objectives and within requirements outlined in contractual agreements.
5. Identifiability of data	Always use the least identifiable data needed to fulfil your task.
6. Data minimization	Only use the minimum but adequate amount of data to accurately carry out your specified purpose.
7. Fitness for purpose	Where possible, users of data need to understand why data was collected, how it has been processed, and any data quality challenges, so they can assess if the data is fit for their purpose, and take account of any data limitations in the interpretation of their analyses.
8. Data storage and retention	Data extracts should be stored securely and only be retained for as long as it is needed, in compliance with contractual conditions. Users of data should not save down multiple extracts of the same data.
9. Data linkage	Data linkage across person-level data assets can increase the risk of identification if not undertaken in a highly governed manner. Before linking data, you should always check (1) if the contract allows you to link data, (2) that you will not increase the risk of identification and (3) that the purpose that you wish to link data are allowed for each of the datasets involved.
10. Disclosure rules	Data assets, countries and contracts may have specific rules for disclosure, these must be understood and adhered to when sharing analysis outputs.
11. Data transit rules	All data must be protected in transit.
12. Data sharing & chain of custody	Understand any restrictions, including country specific rules when sharing data. All data handling responsibilities for data assets must be passed on and explained to the client organizations.
13. Challenge compliance	It is everybody's responsibility to operate in a compliant manner and to highlight non-compliant practice to your line manager or privacy team.

It is important for organisations using data to have a training programme for their data workforce which supports data users to understand their responsibilities when using data, so they are aware of the types of data being used , the risks involved, and how to mitigate them.

[1] The Five Safes was devised by Felix Ritchie at the UK Office for National Statistics (ONS) in 2002 to describe its secure remote-access Virtual Microdata Laboratory (VML); Luk Arbuckle has also written an article about "The 5 Safes of Risk-Based Anonymization". Privacy Analytics White paper, 2020

Innovative data access and uses

Emerging RWD sources, different access and governance models, enabling technologies and sophisticated analytical tools support the emergence of innovative tools/platforms and research project designs.

Below are some examples of such approaches and corresponding case studies:

- **Expanding and leveraging data networks designed to support specific studies** where data may be hard to access. For example, IQVIA's Oncology Evidence Network can be linked to patient outcomes, making comparative effectiveness studies easier to deliver. In addition, some centers in the network are routinely using patient reported outcome tools, supporting quality of life assessments.

Or the IQVIA's Federated Network can be activated to enable large-scale and reproductible research with 300 data partners from 30+ countries

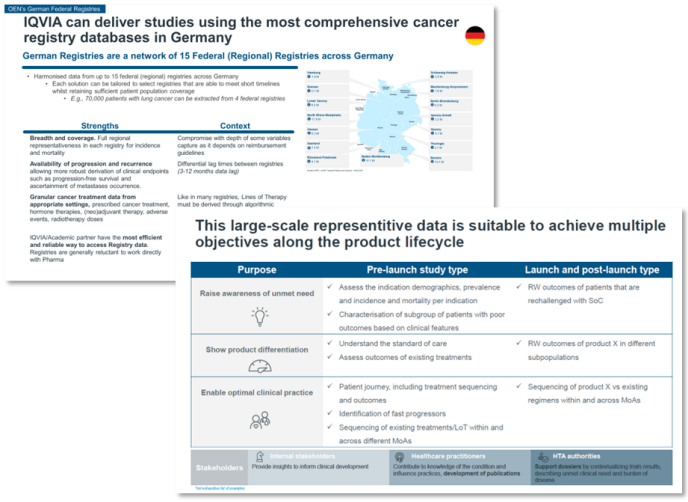
- **Multi country studies:** with increased interoperability starting with the adoption of common data model such as OMOP, multicountry databases studies become increasingly feasible.
- **Enriching data by sourcing additional information** through review of medical records, or adding locally collected data such as patient-reported outcomes
- **AIML** to navigate the increasingly complex healthcare ecosystem with precision, achieve results with greater **speed**, and **scale** to meet varying needs. With AIML platforms for example, patient pathways can be deciphered, unmet needs identified thus supporting healthcare systems and management improvement.

Expanding and leveraging data networks

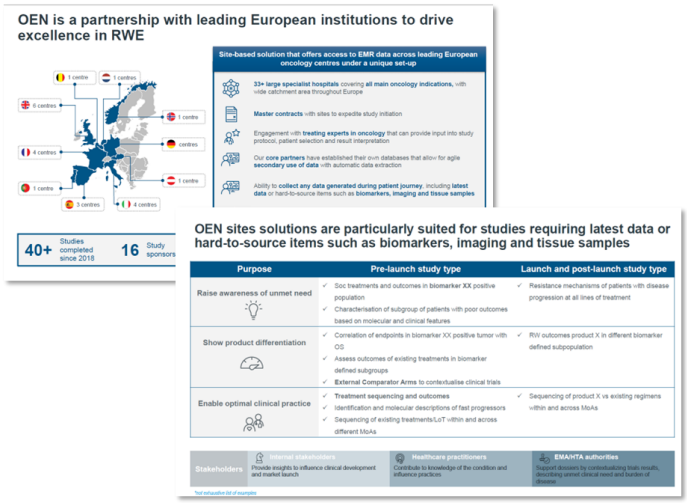
The next pages detail some case studies or use cases illustrating the above examples.

ONCOLOGY EVIDENCE NETWORK

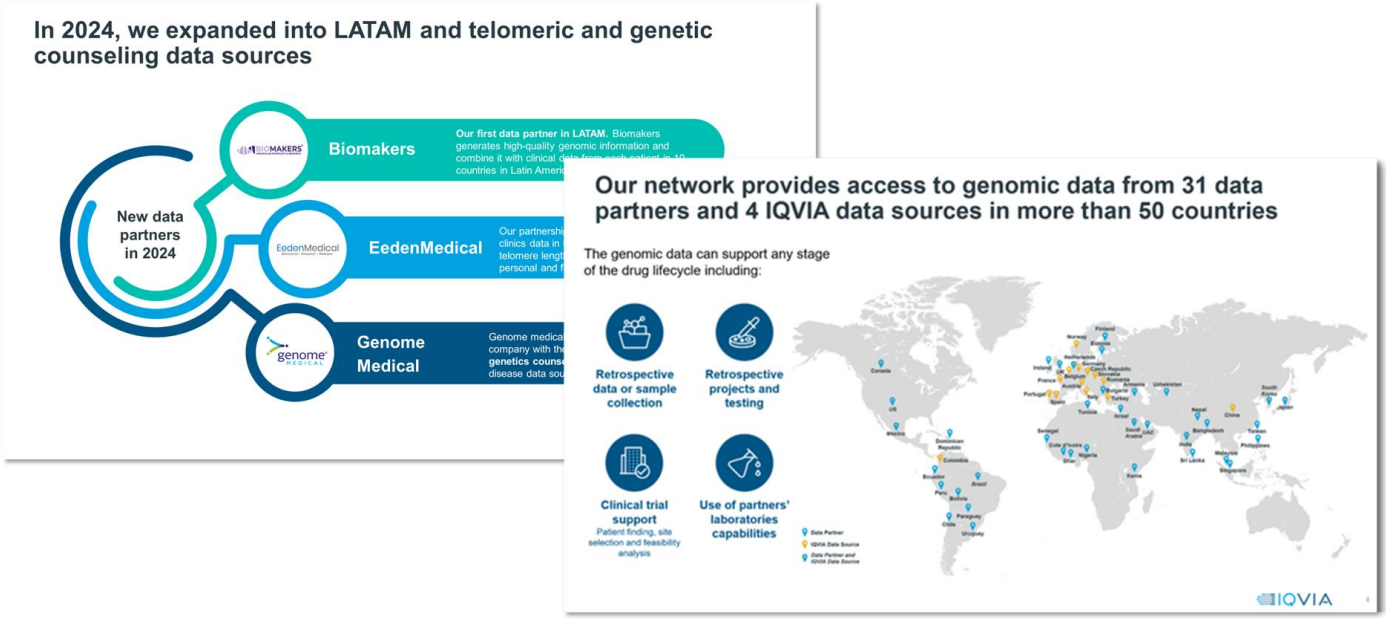
OEN Registries solutions
Focus on German Federal Registries



OEN Sites solutions



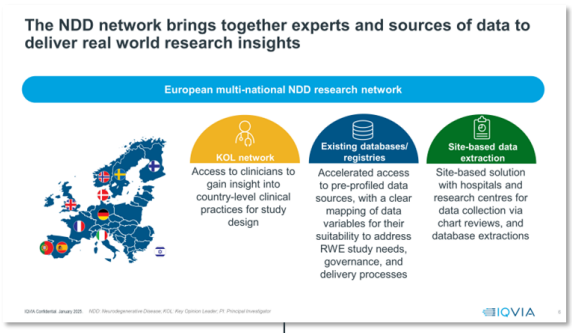
PRECISION MEDICINE / GENOMICS

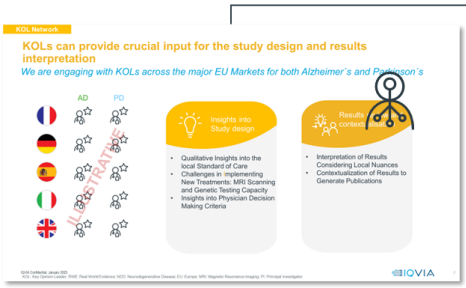


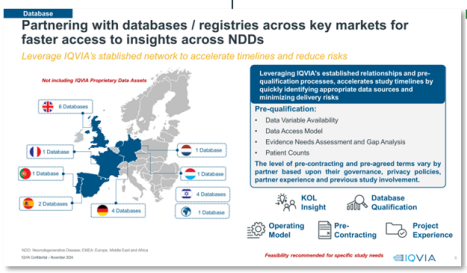
EMEA NEURODEGENERATIVE DISEASES (NDD) NETWORK

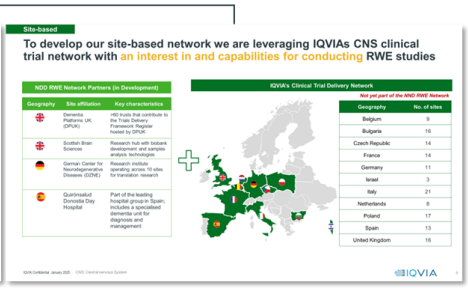
ⓘ Current focus: Alzheimer's and Parkinson's diseases

Progressively expanding to other conditions as the network grows









Attrition rates from first- to third-line therapy in HER2+ metastatic breast cancer in Europe

San Antonio Breast Cancer Symposium®, December 5–9, 2023 - Poster PO3-16-11

AUTHOR(S)

Peter S Hall,[1,2] Sue Cheeseman,[3] Paul Cottu,[4] Markus Krebs,[5] Christian W Scholz,[6] Emilio Bria,[7] Armando Orlandi,[7] Nuria Ribelles,[8] Mahéva Vallet,[1,2] Nicolas Niklas,[9] Catherine Hogg,[10] Markus Lucerna,[11] Simon M Collin,[12] Amanda Logue,[13] Gráinne Long[12]

OBJECTIVES

- Understand and characterize contemporary attrition rates of patients who completed a line of therapy [LOT] but did not receive the subsequent LOT) after first-line (1L) and second-line (2L) therapy among women with human epidermal growth factor receptor 2-positive (HER2+) metastatic breast cancer (mBC) receiving routine care at selected oncology centers in (France, Germany, Italy, Spain) and the UK.
- Highlight the potential lost opportunity for patients with HER2+ mBC to benefit from optimal targeted treatments at the earliest indicated LOT.

INTRODUCTION

- Women with HER2+ mBC who experience progression after 1L therapy (trastuzumab, pertuzumab, and a taxane) typically require a subsequent LOT1.
- Trastuzumab emtansine (T-DM1) was standard of care in the 2L setting until the recent approval (July 2022 in the EU) of trastuzumab deruxtecan (T-DXd) after ≥1 anti-HER2 regimen2.
- Receiving optimal targeted therapy in the earliest indicated setting is important to maximize the likelihood and durability of a clinical benefit.
- As new therapies become available, understanding treatment patterns by LOT may help guide treatment decision making and inform the optimal treatment paradigm for patients with HER2+ mBC.

METHODS

- In this ongoing multicenter observational study, electronic medical record (EMR) data were collected retrospectively from women ≥18 years old diagnosed with HER2+ mBC between 2017 and 2021, in EU4 countries and the UK.
- Structured EMR data and manually abstracted unstructured data from oncology centers were curated. Patients had the opportunity to be followed up for ≥12 months from mBC diagnosis.

RESULTS AND INTERPRETATION

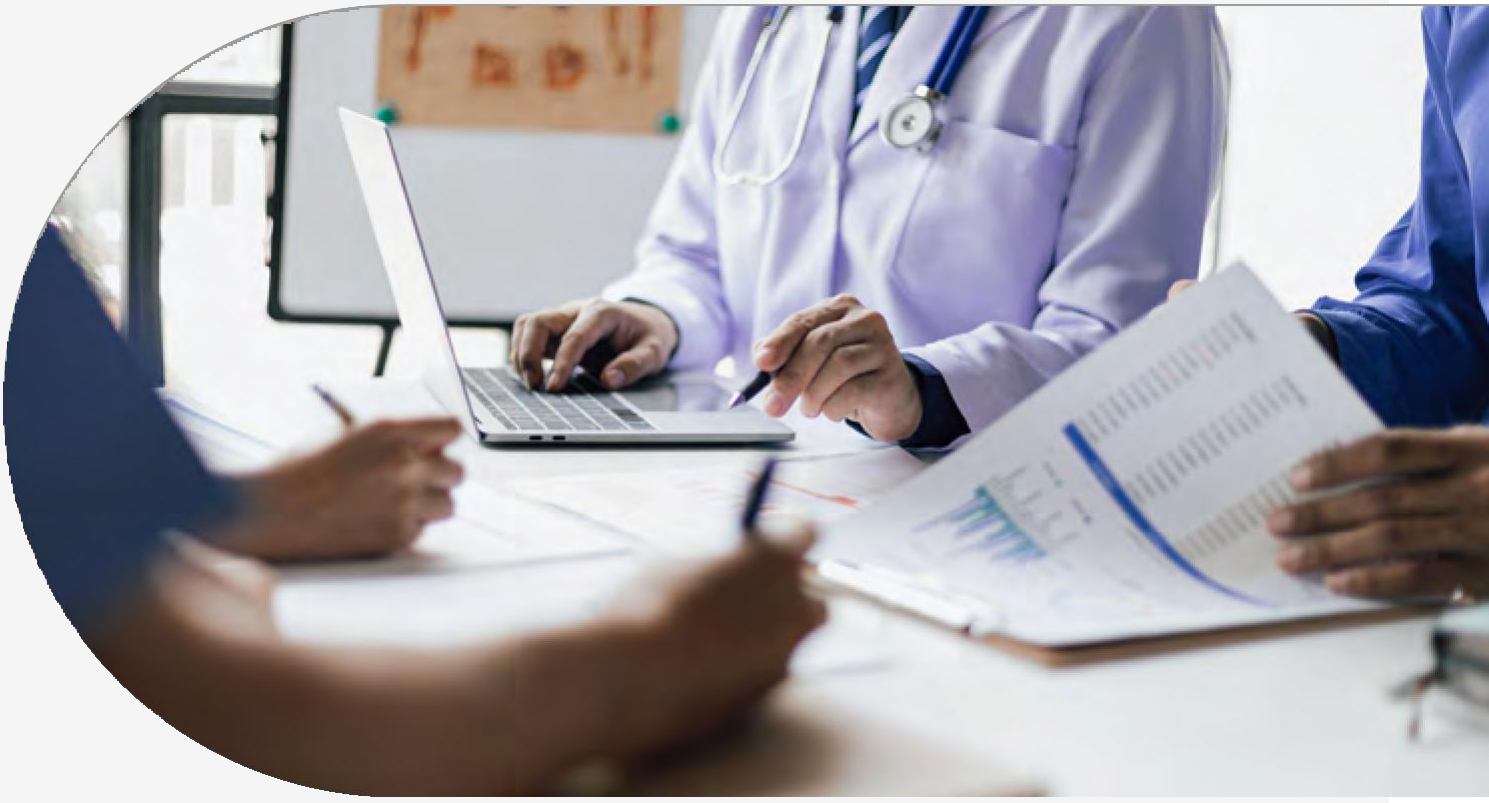
- This analysis included data from 496 women across seven sites in five countries: 7 sites in the 5 above mentioned countries.
- Overall median duration of FU was 41.1 months (95% confidence interval [CI] 22.4, 52.8). A total of 59.1% of patients were postmenopausal, 60.9% had de-novo disease, 60.9% had Stage IV disease, and 27.4% had ≥4 metastatic sites. Overall, 41.9%, 26.4%, and 12.3% received a total number of one, two, and three LOTs per patient, respectively.

CONCLUSIONS

- These data from routine clinical practice indicate that 29.6% and 34.2% of patients with HER2+ mBC who completed 1L and 2L therapy, respectively, did not receive a subsequent LOT, which was primarily due to death, move to end-of-life palliative care, loss to follow up (FU), and 'other' reasons for attrition; this highlights the importance of using optimal HER2-directed treatments in the earliest indicated setting.
- Further research is needed to understand the high variation in reasons for attrition, as this may reflect differences in patient fitness, clinician and patient attitudes to treatment, and access to treatments across centers.
- Future analyses for this study will focus on characterizing treatment patterns and real-world outcomes.

DATABASE SOURCE Data were extracted from de-identified, coded and uncoded hospital electronic patient records across seven sites in France, Germany, Italy, Spain, and UK.

[1] University of Edinburgh, UK; [2] Edinburgh Cancer Centre, Western General Hospital, NHS Lothian, UK; [3] Leeds Teaching Hospitals NHS Trust, UK; [4] Department of Medical Oncology, Institut Curie, Paris, France; [5] Comprehensive Cancer Center Main franken, University Hospital Würzburg, Germany; [6] Department of Hematology and Oncology, Vivantes Klinikum Am Urban, Berlin, Germany; [7] Agostino Gemelli University Policlinic, Rome, Italy; [8] Hospital Universitario Virgen de la Victoria de Málaga, Spain; [9] IQVIA, Frankfurt, Germany; [10] IQVIA, London, UK; [11] Daiichi Sankyo Europe GmbH, Munich, Germany; [12] Global Medical Affairs, Evidence Generation & External Alliances, Oncology Outcomes Research, OBU Medical, AstraZeneca, Cambridge, UK; [13] Global Medical Affairs, Regional Medical Affairs, OBU Medical, AstraZeneca, Cambridge, UK.



IQVIA EMR DATA, REAL WORLD EVIDENCE

Characterising people with focal drug-resistant epilepsy: A retrospective cohort study

Epilepsy & Behavior, Volume 149, 109540, December 2023

AUTHOR(S)
Caroline Benoist, Simona Boccaletti, John Paul Leach, Agnese Cattaneo, Anna Chaplin, Luis Antunes, Franca Heiman, Josemir W. Sander

HIGHLIGHTS

- Real-world data of 1,075 individuals with F-DRE in 6 European countries.
- The analysis covered six years.
- Levetiracetam is the most prescribed first line in all countries but the UK.
- Depression and anxiety are the most frequently reported psychiatric disorders.
- Specialist/GP consultation is sought mainly post F-DRE than post epilepsy diagnosis.

STUDY OBJECTIVES

To describe the demographics, clinical characteristics, drug treatment outcomes, healthcare resource utilization, and injuries among people with focal drug-resistant epilepsy (F-DRE) analysed separately for six European countries.

METHODS

We used electronic medical record data from six European (Belgium, Spain, Italy, France, UK and Germany) primary care/specialist care databases to identify antiseizure medication (ASM) treatment-naïve people (aged ≥ 18 years at F-DRE diagnosis). They were followed from their epilepsy diagnosis until death, the date of last record available, or study end. We used descriptive analyses to characterise the F-DRE cohort, and results were reported by country.

CONCLUSIONS

We attempted to understand better the burden of illness and treatment patterns of people with F-DRE. Our results show no one-size-fits-all approach in the appropriate selection of ASM, and the drug's potential risks and possible benefits must be individually considered. Identifying and considering comorbidities must be an integral part of the management and should influence ASM choice. Our findings may generate valuable information on actual treatment practices and features of people with F-DRE at the primary and specialist care levels, which may support future treatment recommendations and improvements in clinical care.

SIGNIFICANCE

No one ASM is optimal for all people with F-DRE, and the risks and benefits of the ASM must be considered. Comorbidities must be an integral part of the management strategy and drive the choice of drugs.

DATABASE SOURCE Data were extracted from proprietary IQVIA EMR in UK, Germany, France, Spain Italy and Belgium.

Enriching data by sourcing additional information

PATIENT-MEDIATED DATA ACCESS

- Expanding reach of data collection through patient-provided access to EMR and claims
- Opportunities for long-term follow-up and direct-to-patient studies

EMR TO EDC (E2E)

- Leveraging interoperability standards to access EMR for EDC pre-population
- Reducing data entry burden at sites
- Maximizes efficiency and data quality
- Opportunities for tech-enabled chart reviews

COMBINING PRIMARY AND SECONDARY DATA

- Linking multiple data sources at the patient level
- Leveraging secondary data from EMRs, claims, registries and research cohorts
- Capabilities to integrate data from novel sources such as smart devices
- Generating high-quality, longitudinal patient data

DATA ENRICHING OPTIONS INCREASE AS LINKAGE CAPABILITIES EXPAND

Country	Available Data Sources	Data Collection Modalities
UK	EMR – Primary Care	Patient mediated, data enabled sites, Enriched
	Specialty EMR – Ophthalmology	
	Registry – Oncology	
	Mortality Data	
US	Claims – Primary & Secondary Care	Patient mediated, data enabled sites, Enriched
	EMR – Primary & Secondary Care	
	Specialty EMR – Oncology	
	Registry – Mortality Data	
Canada	EMR – Primary Care with Specialist Care	Enriched
Germany	EMR – Primary Care with Specialist Care	Enriched
	EMR – Primary Care with Secondary Care enabled	
France	EMR – Primary Care with Specialist Care	Enriched
	Claims – Primary & Secondary Care	
Denmark	EMR – Primary & Secondary Care	Enriched
Sweden	EMR – Primary & Secondary Care	Enriched
	Registries	
Norway	EMR – Primary & Secondary Care	Enriched
	Registries	
Finland	EMR – Primary & Secondary Care	Enriched
	Registries	
Netherlands	EMR – Primary Care with Secondary Care enabled	Enriched
	Registry	
Belgium	EMR – Primary Care with Secondary Care enabled	Enriched
Brazil	Claims – Secondary & Tertiary Care	Enriched

Data enabled sites have been integrated with technology to automatically pull EMR data to fill a CRF in the EDC

REAL WORLD EVIDENCE

An enriched case study: COVID Active Research Experience (CARE) linked to claims data

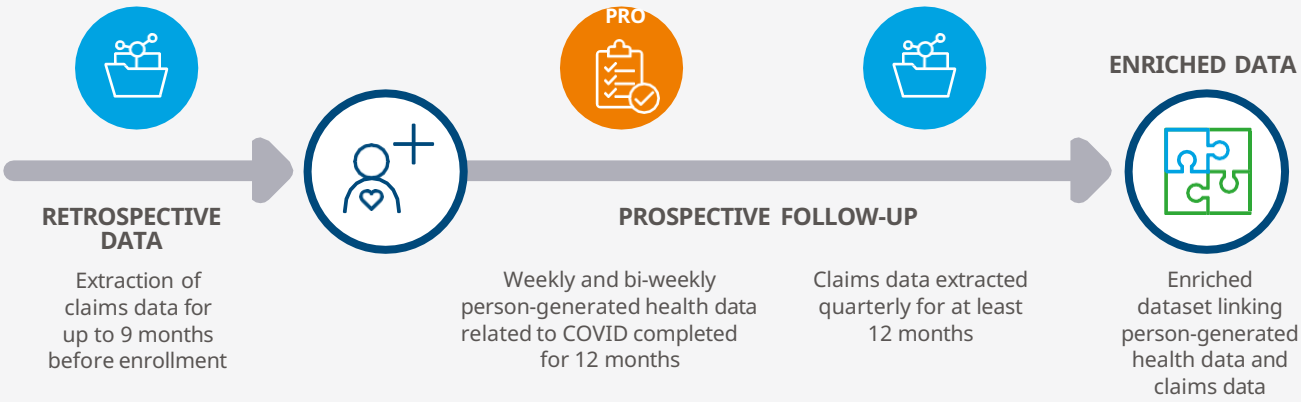
SITUATION

- IQVIA is contributing to a global health effort to learn more about COVID-19, and its treatments and vaccines.
- The CARE project is focused on describing how COVID-19 infections, treatments, and vaccines affect people in the community.

STATUS

Data collection is ongoing

SOLUTION



- The CARE registry of community participants is focused on symptom occurrence, severity, and progression or resolution over time related to the SARS-CoV-2 virus.
- Participants consent to participate in questionnaires delivered via a secure, web-based portal and to allow linkage of this person-generated health data with medical and pharmacy claims data.
- Person-generated health data are linked to claims data through deterministic linkage using proprietary algorithms and a trusted third party to protect participant identities and comply with patient data privacy regulations.
- The enriched methodology utilized for CARE allows for a flexible approach to evidence generation on the ever-changing questions regarding the epidemiology of the COVID-19 pandemic and patients' real-world experience.

DATABASE SOURCE Proprietary IQVIA US medical and pharmacy claims data

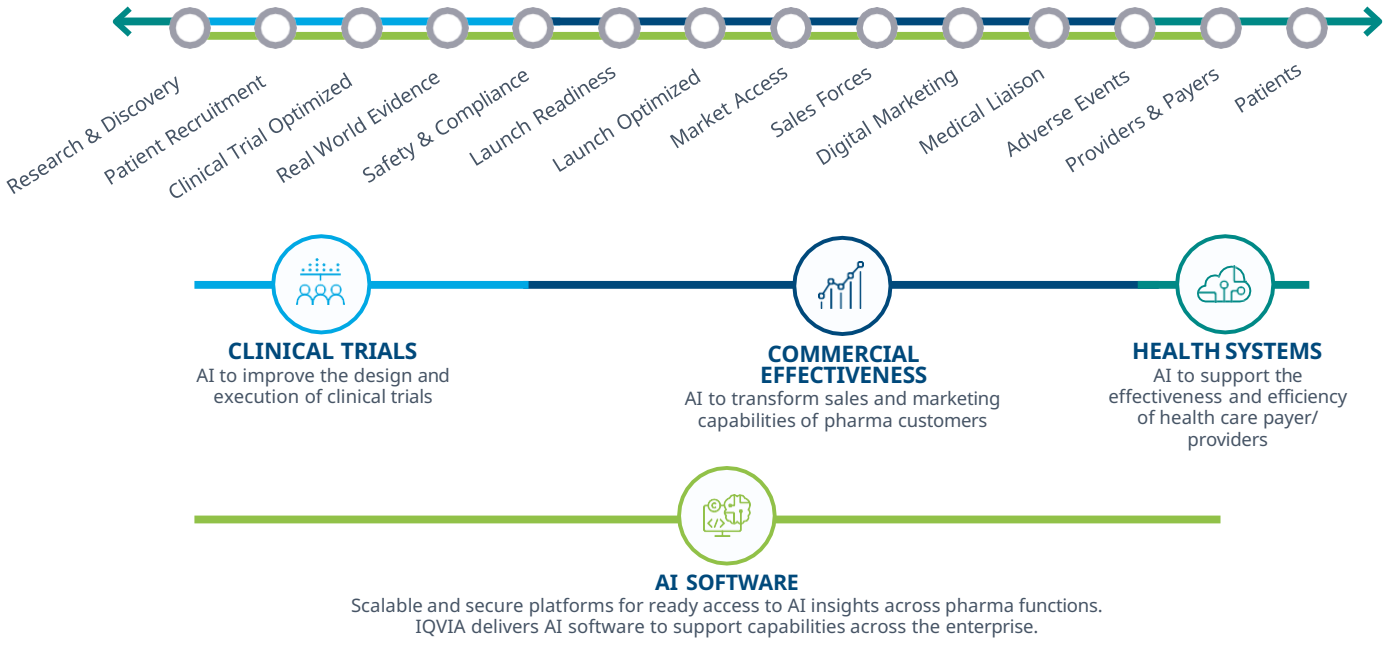
Artificial intelligence and machine learning

A VIEW OF OVERALL AI EVOLUTION & EXAMPLES OF HEALTH DATA ANALYTICS

What is the technique?	Brief description	Good for / not good for
Advanced analytics (statistical operations in pipelines)	Statistical operations within data processing pipelines. To uncover valuable insights, patterns and correlations, facilitating data-driven decision-making.	Historical data analysis, predictive modeling, & optimizing business processes Real-time decision-making or analysis of constantly changing data
AI (probabilistic engines using ML or complex ontologies & graph)	Systems powered by probabilistic engines, machine learning algorithms and complex knowledge representations (ontologies & graphs). Mimic humanlike reasoning and understanding in machines.	Tasks automation, complex decision making, NLP, computer vision & enhancing UX Tasks needing deep understanding of abstract concepts beyond pattern recognition
Machine learning (recursive training on datasets)	An approach where models are trained iteratively on datasets to learn patterns and make predictions. It involves various algorithms to recognize and generalize from data.	Image recognition, recommendation systems, and predictive modeling Not for extremely limited or noisy data as model performance depends on data quality.
Transformers (neural network based machine learning)	Class of neural network models renowned for their sequence-to-sequence capabilities & ability to capture contextual relationships in data. They utilize self-attention mechanisms to process data, making them ideal for various natural language processing tasks.	Language translation, text summarization, and sentiment analysis Less efficient for tasks with straightforward solutions achievable with simpler algorithms
Language models (retaining knowledge of text in the model)	LMs are large neural networks designed to understand and generate text. They are pretrained on vast text corpora and retain knowledge to provide coherent, context a ware text generation.	Chatbots & question answering systems, also store textual knowledge for reference) Not for physical actions or decision making outside the generation of text-based answers
GPT (a specific probabilistic word generator using LLMs)	Specific type of LM that excels in generating coherent and contextually relevant text. It uses a vast pretrained dataset to generate word sequences based on learned patterns but without inherent correctness or ethical judgment.	Content generation, text completion, creative writing assistance, & answering text-based queries. Not for ethical or moral decision making,
HGT, RetNet, ...	**GPT is not the last word in AI. HGT (Hyper Graph Transformer) maybe be better for ontology data, and RetNet runs LLMs x16 faster than transformers**	

APPLYING IQVIA AI SOLUTIONS ACROSS THE PRODUCT LIFE CYCLE AND BEYOND

AIML/NLP solutions unleash tangible value, bringing precision, speed and scale to the entire product life-cycle, from molecule to market as per IQVIA's use case examples below:



USE CASE EXAMPLES: BUILDING A MODULAR ANALYTICS PLATFORM AS CONSOLIDATED SOURCE OF MEDICAL AND SCIENTIFIC REAL-WORLD ANALYTICS IN ASTHMA AND COPD

AI/ML apps as integration layer for multiple real-world analytics use cases

Background & Client Vision

- Client is working towards **transforming the lives of asthma/COPD** patients by addressing current gaps across the spectrum of the condition
- Client follows **multidisciplinary strategy**, including **commercial and medical** field and scientific activities.
- Scale earlier pilot work with IQVIA into a **data platform** that serves as **single source of real-world analytics**

Outcome

- Built **analytics platform** and **data pipeline** that feeds client's **quarterly market** and disease monitoring
- Supported four medical **spin-off studies** and **publication** based on AI/ML apps
- **Renewed** and **expanded scope** to include and monitor **spin-off analytics** KPIs

Stakeholders

Global
- Medical Program Director
- Policy Director
- Assoc. Dir. Epidemiology

Sponsor complex add on KPIs

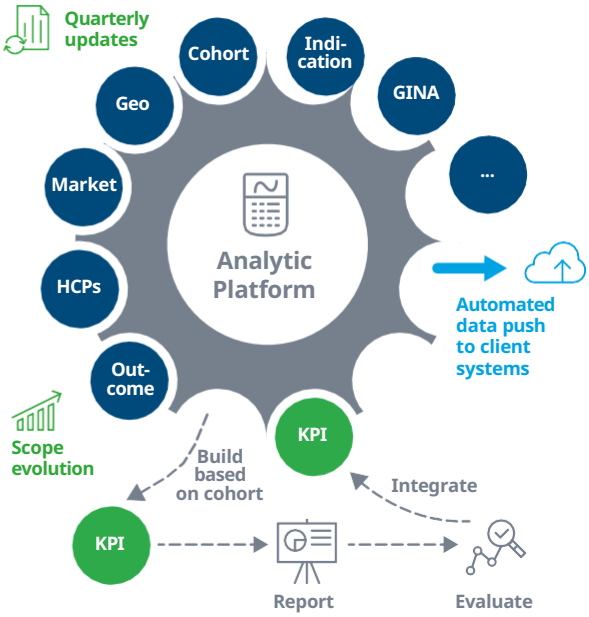
Affiliate, Medical
- Scientific Advisor
- MSL Manager COPD
- MSL Manager Asthma
- Med Strategy Manager

Define and consume KPIs in spin off projects

Affiliate, Data Science
- Dir Data Science
- Sr Data Scientist

Build dashboard from quarterly updates. Coordinate needs with med teams

Engagement Lifecycle



PFIZER + IQVIA ABSTRACT ACCEPTED FOR IDWEEK — POSTER IN PROGRESS

Use of natural language processing to extract published real world data on a covid vaccine and antiviral treatment

"The COVID-19 global pandemic generated an exponential increase in scientific publications over the last 3 years. LitCovid includes ~350,000 articles related to COVID-19, vaccine and antiviral medicine development…"

"We used a two-step approach comprised of (1) an automated NLP system with machine learning and rule-based methods to extract medical data from articles and (2) a manual evidence synthesis approach to confirm accuracy by expert review to create a RWD dataset for effectiveness related to BNT162b2 (Pfizer-BioNTech COVID-19 Vaccine) and nirmatrelvir/ritonavir (Paxlovid)"

"Elapsed NLP-based extraction time was compared to manual extraction by 6 scientific experts over a 10 day test, and speed gains of 2.5-3.0x were achieved."

Prepared by IQVIA
Anna Ssentongo, DrPH • Lenon Mendes Pereira, PhD, Yuting Kuang, PhD • Jennifer Uyei, PhD, MPH

IDWeek

AI GOVERNANCE PRINCIPLES

Fairness

- Adjusting **known bias in training** of specific AI engine (e.g., CONAN)
- Profiling all training data and potential bias (e.g., **health data catalog**)

Respect

- Building **privacy assured "green zones"** for ML experimentation
- Systematic analytics of re-identification risk **via dedicated AI software**

Transparency

- Publishing an inventory of **peer-reviewed methods in AI**
- **Benchmarking AI** against alternatives

Accountability

- Mandated AI Council, including a robust AI operating model *(now a GPT council too)*
- Does our AI Work? **AI claims & ROI**, prospective validation for patient outcomes

Generate trust and scientific validity in AI use

LEARN MORE: Download a copy of our white paper with case studies *Pragmatic Application of Healthcare AI Governance* - click here.

Glossary

AIML Artificial Intelligence & Machine Learning.

APAC Asia Pacific.

BMI Body Mass Index.

CPGs Consumer Packaged Goods. Items used daily by average consumers that require routine replacement or replenishment.

CPRD Clinical Practice Research Datalink. Observational and interventional real-world research service operating as part of the UK Department of Health. (Find more information here)

CV Cardio-Vascular. Referring to the circulatory system, which consists of the heart and the pulmonary and systemic circulation, which transport nutrients and oxygen to and remove waste products and carbon dioxide from organs and tissues.

DB Database.

DUS Drug Utilization Study. The study is aimed to answer questions such as (i) how much is used of certain drugs in a country, region, hospital or in a primary care setting, (ii) what is the drug used for (indication), (iii) how is it used (dose and duration of treatment), and (iv) who are using it (age, gender)?

eCOA Electronic Clinical Outcome Assessment.

EHDEN European Health Data and Evidence Network.

EMA European Medicines Agency. The European Medicines Agency is a decentralized agency of the European Union, located in London. The Agency is responsible for the scientific evaluation of medicines developed by pharmaceutical companies for use in the European Union. It began operating in 1995.

EMR Electronic Medical Records. A digital version of patients' chart comprising the data gathered by a clinician's office. It contains the patient's medical and treatment history from one practice.

EHR Electronic Health Records. EHR contains the patient's records from multiple doctors and provides a more holistic, long-term view of a patient's health.

ER Emergency Room.

FDA Food & Drug Administration. An agency of the U.S. Department of Health and Human Service. The FDA is responsible for assuring the safety, efficacy and security of human and veterinary drugs, biological products, medical devices, and other products.

GDPR Global Data Protection Regulation (EU 2016/679). A legal framework that sets guidelines for the collection and processing or personal information from individuals living in the European Union.

GPs General Practitioners.

HCC Hepatocellular carcinoma (type of primary liver cancer).

Glossary – *continued*

HCRU Health Care Resource Utilization.

HEOR Health Economics & Outcomes Research.

HES Hospital Episode Statistics. A data warehouse containing details of all admissions, outpatient appointments and A&E attendances at NHS hospitals in England.

HIPPA Health Insurance Portability and Accountability Act. A federal act, passed in 1996 with the primary goals to make it easier for people to keep health insurance, protect the confidentiality and security of healthcare information and help the healthcare industry control administrative costs. Federal law regarding privacy of patient health information.

HMSL Hospital Marketing Services Ltd is a Patient Diary Study collecting summary patient information about the treatment by the secondary care physician.

HTI Hospital Treatment Insights.

ICD International Classification of Disease.

ICU Intensive Care Unit.

IMRD IQVIA Medical Research Data. A large database of anonymized electronic medical records collected at primary care clinics throughout the UK.

IRB Institutional Review Board. An administrative body established to protect the rights and welfare of human research subjects recruited to participate in research activities conducted under the auspices of the institution with which it is affiliated.

LOT Length of Treatment; Line of Treatment (the order in which different therapies are given to people as their disease progresses i.e., first line of treatment, second line of treatment, etc.).

LPD Longitudinal Patient Data.

LRx Longitudinal Prescriptions. Longitudinal patient prescription dataset based on retail pharmacy data and coding centers. It enables the longitudinal tracking of patient prescription activity.

Market Sizing Defined as estimating the number of buyers of a particular product, or users of a service.

MBDS Minimum Basic Dataset (designed and used by general practitioner).

MS Multiple Sclerosis.

NBRx New to Brand Prescription. A new medication prescribed by a physician to a patient.

NGS Next Generation Sequencing.

NHI National Health Insurance.

Glossary – *continued*

NLP Natural Language Processing. Natural language processing is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language.

NRx New Prescription. Number of new prescriptions (NRx; New Prescriptions) written by physicians for a particular drug over a specific period of time. NRx does not include prescriptions for refills, but does include renewals, which are scripts that the patients get when they run out of refills. In contrast, TRx (Total Prescriptions) includes refills AND renewals.

NSCLC Non-small cell lung cancer.

OHDSI Observational Health Data Sciences and Informatics. It is a public initiative independent from IQVIA, enabling faster, more reliable studies and/or expanded data access via the OHDSI global networks.

OMOP Observational Medical Outcomes Partnership. The Observational Medical Outcomes Partnership (OMOP) is a public-private partnership established to inform the appropriate use of observational healthcare databases for studying the effects of medical products.

OMOP CDM OMOP Common Data Model (CDM) transforms observational health data into a structured, common format.

OTC Over the counter. OTC drugs which are legally allowed to be sold by pharmacists without need for a prescription.

PASS Post Authorization Safety Study.

PROs Patient Reported Outcome. Health outcome directly reported by the patient who experienced it, contrary to an outcome reported by someone else, such as a physician-reported outcome

PV Pharmacovigilance. Science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other medicine/vaccine related problem.

QoL Quality of Life. Perceived quality of an individual's daily life, that is, an assessment of their well-being or lack thereof.

RA Rheumatoid Arthritis.

RA/MS Rheumatoid Arthritis/ Multiple Sclerosis.

REAL/ OEN A research network of Europe's larger cancer hospitals working with IQVIA to provide high-quality, real-world data reflecting the latest clinical practice.

RWD Real World Data is data derived from real-world settings such as electronic health records, health insurance claims, hospitals, etc.

Glossary – *continued*

RWE Real World Evidence is evidence derived from analysis of RWD for example, clinical outcome measured for patients in real-world setting.

TNM A system to describe the amount and spread of cancer in a patient's body, using TNM. T describes the size of the tumor and any spread of cancer into nearby tissue; N describes spread of cancer to nearby lymph nodes; and M describes metastasis (spread of cancer to other parts of the body).

TTP Third Trusted Party. An entity other than the owner and verifier that is trusted by the owner, the verifier or both to provide certain services.

CONTACT US

RWDassets@iqvia.com

www.iqvia.com/solutions/real-world-evidence/real-world-data-and-insights