What is data quality, and how data types differ in terms of data quality?

Massoud Toussi, MD, PhD, MBA
Global Category Lead, Evidence from Secondary Data, IQVIA

BBS/BES Seminar
Wednesday 15th March 2022 from 14:30-17:00 CET
Virtual meeting

How well your data meets your needs?

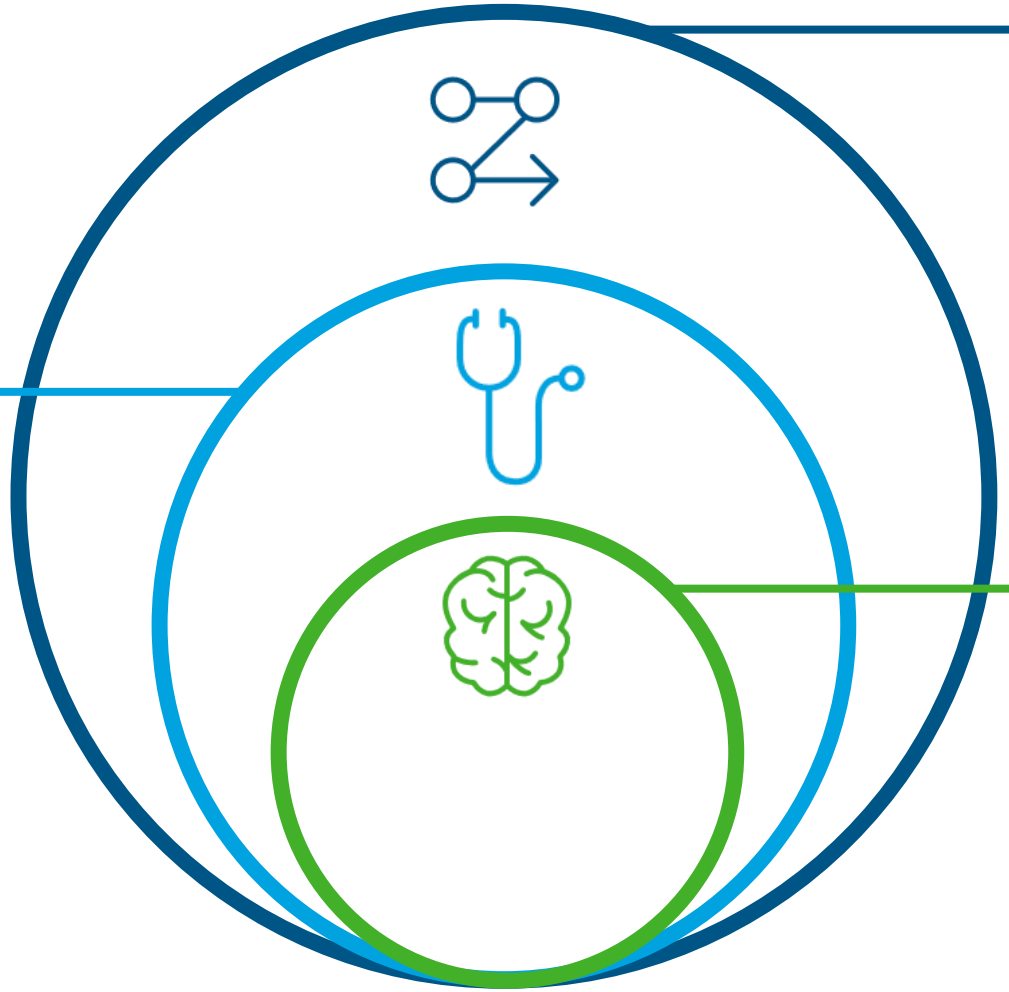
Data quality can be measured at different levels

Database

The storage structure which holds the data.

Data collection

The process of gathering and registering the data

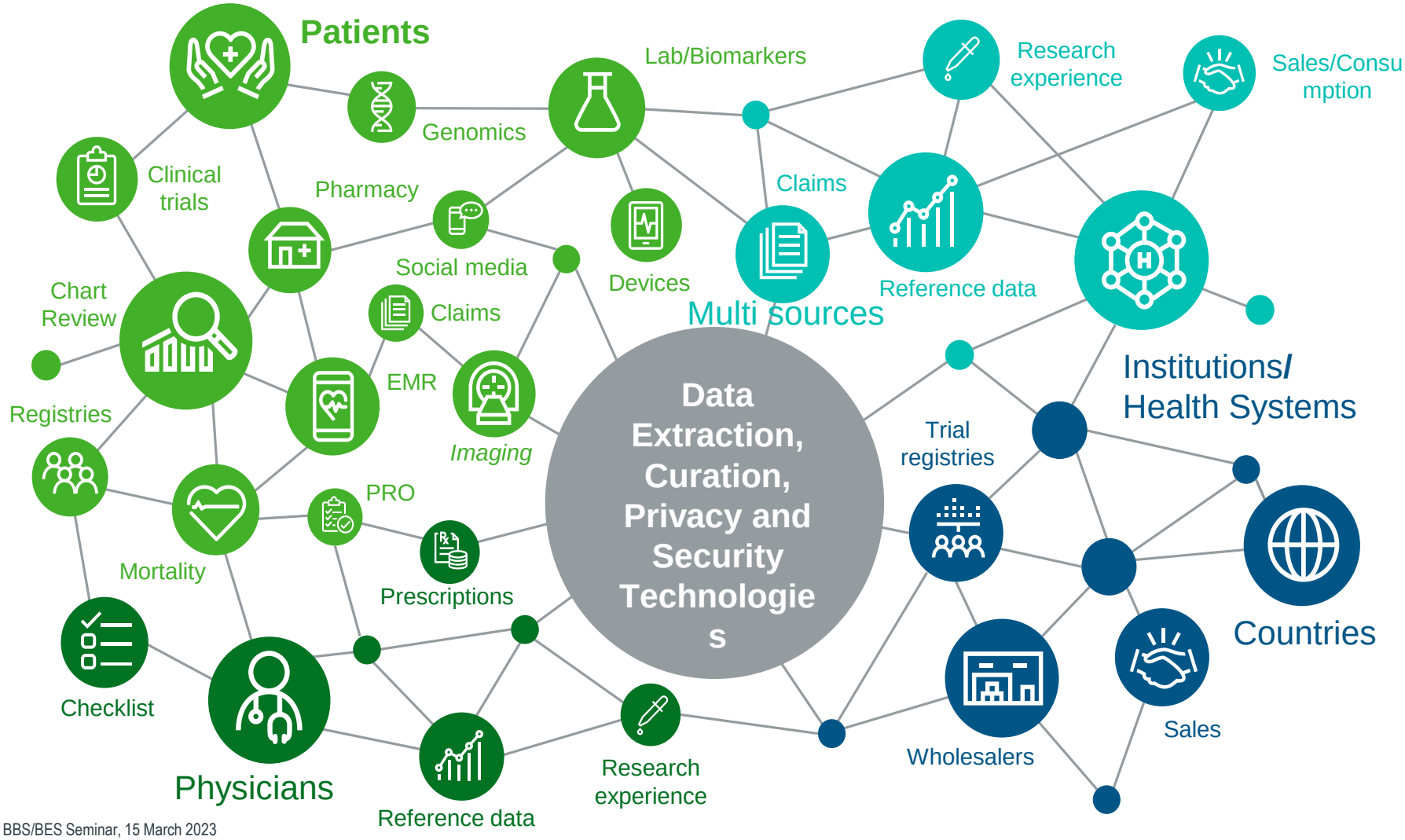


Data

- Factual information which is used for analysis.

The FAIR principle

The ultimate objective of a database should be its re-usability.

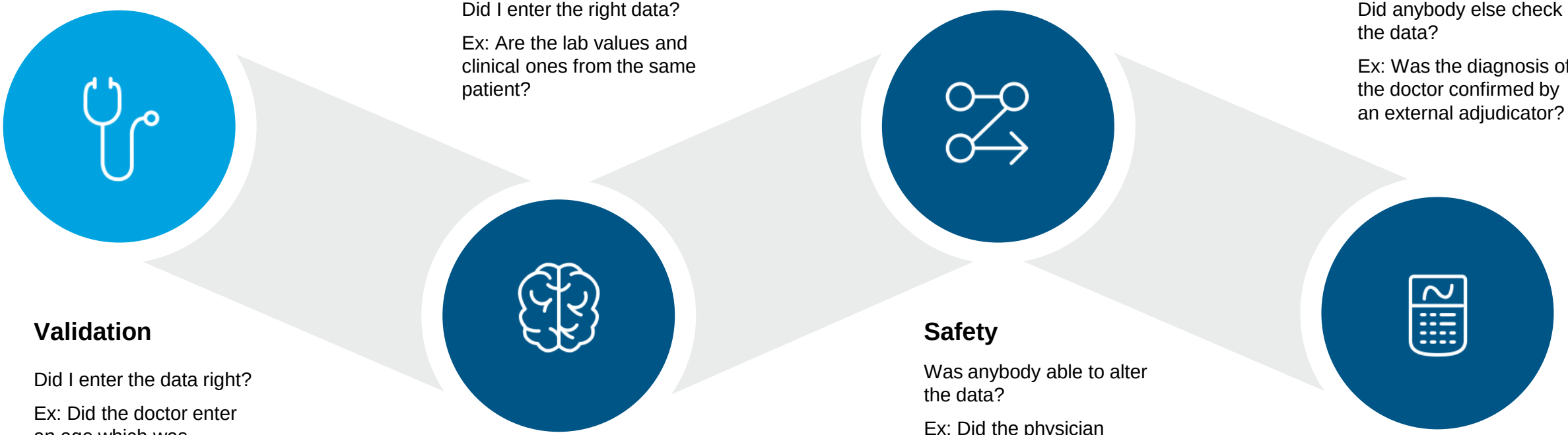


F.A.I.R

- **Findable**
- **Accessible**
- **Interoperable**
- **Re Usable**

Validation, verification, safety, control

Transparency and traceability allow the measurement of these elements.



Validation

Did I enter the data right?

Ex: Did the doctor enter an age which was between 0-140 years?

Verification

Did I enter the right data?

Ex: Are the lab values and clinical ones from the same patient?

Safety

Was anybody able to alter the data?

Ex: Did the physician share their user/password with someone else?

Control

Did anybody else check the data?

Ex: Was the diagnosis of the doctor confirmed by an external adjudicator?

Fitness for purpose

The most important element of fitness-for-purpose is relevance.

Relevance
Does it have the variables I need?

Comleteness
Are all variables filled with values?

Accuracy
Does it have errors?

Linkability
Can I link it to other data sources?

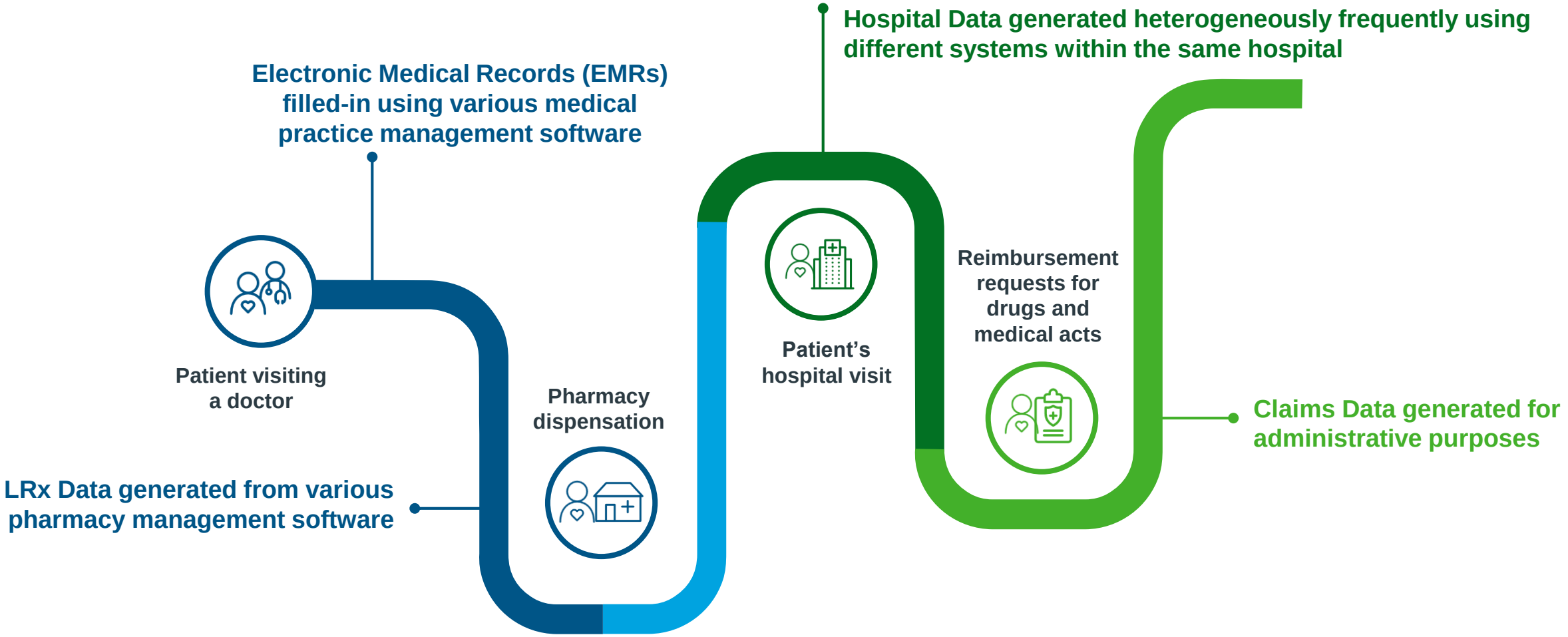
Timeliness
Can I access the data within my deadlines?

Representativeness
Is the population similar to the one I want to study?

Accessibility
Can I access the data at all?

Focus on main sources of secondary longitudinal patient data

Data is generated for administrative purposes not for analytical ones



Claims data

Initial purpose

- Economic management
- Reimbursement

Content

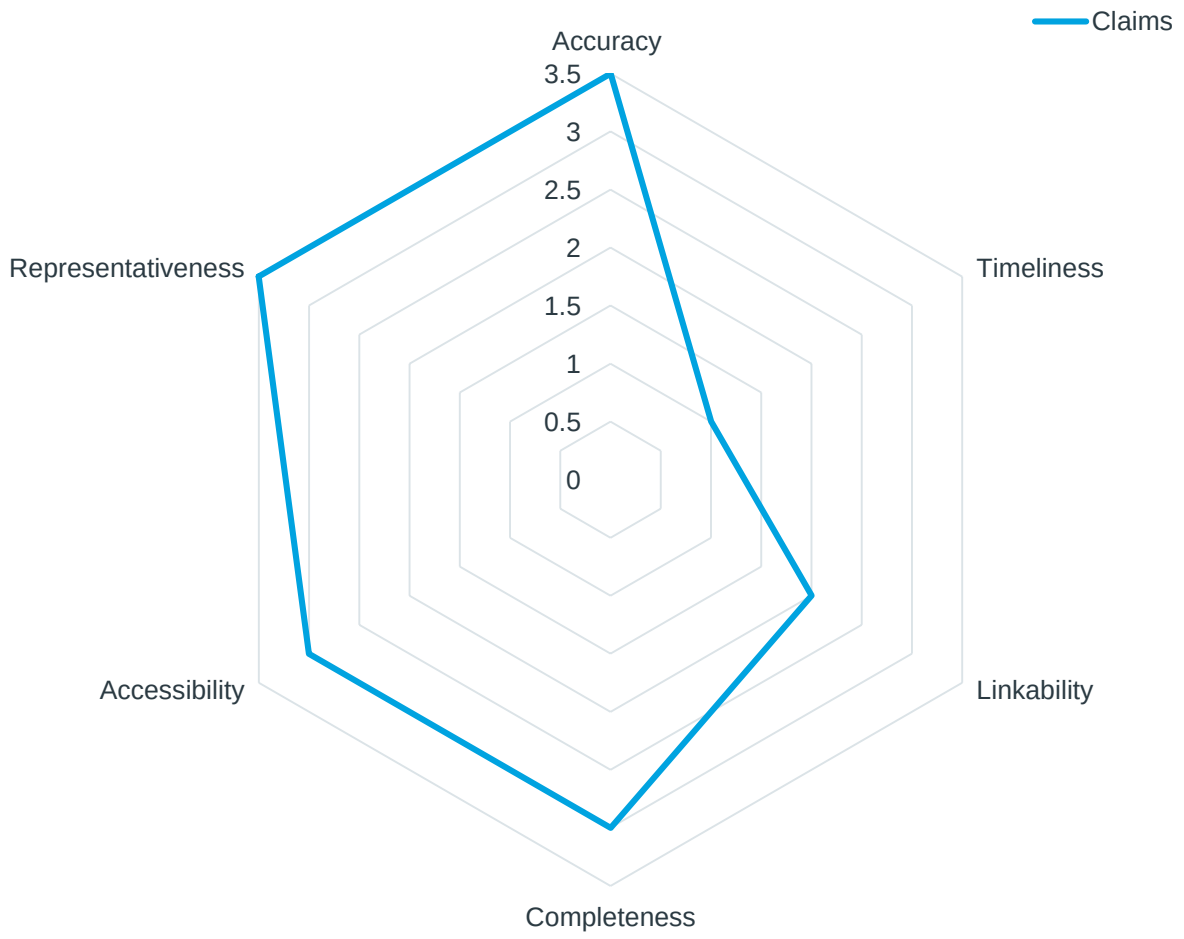
- Demographics
- Diagnoses
- Diagnostic related groups
- Procedures
- Reimbursed drugs/devices

Settings

- Mainly Hospitals
- Increasingly linked to outpatient claims

Good for

- Economic and resource utilization
- Epidemiology
- Healthcare system



EMR data

Initial purpose

- Clinical management
- Patient follow up

Content

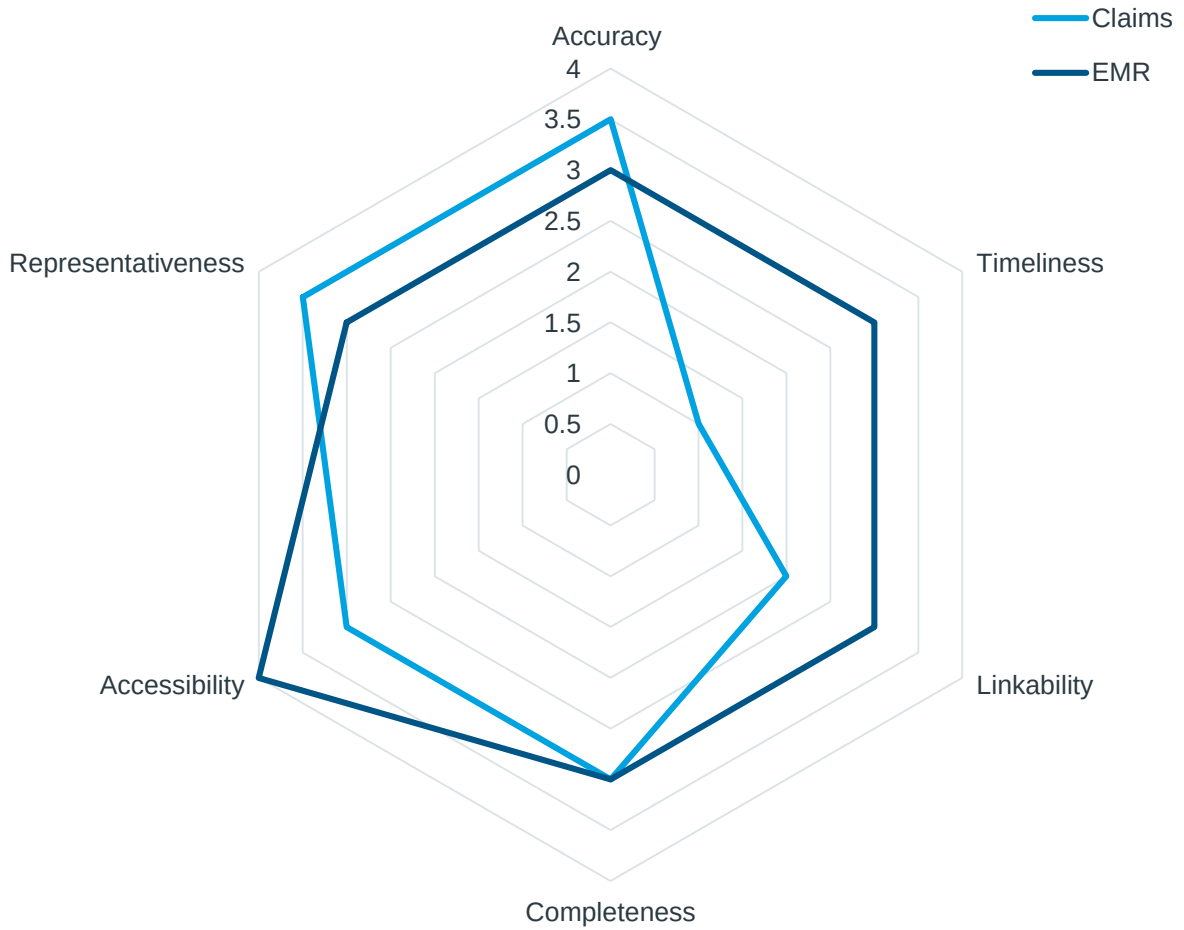
- Demographics
- Diagnoses
- Signs and symptoms, allergies, smoking
- Lab values
- Drugs and to a less extent procedures

Settings

- Mainly primary care
- Increasingly secondary care and hospitals

Good for

- Exposure evaluation
- Drug utilization
- Disease epidemiology
- Benefit-risk assessment
- Unmet needs, burden, adherence



Pharmacy records/sales data

Initial purpose

- Sales management
- Benchmarking

Content

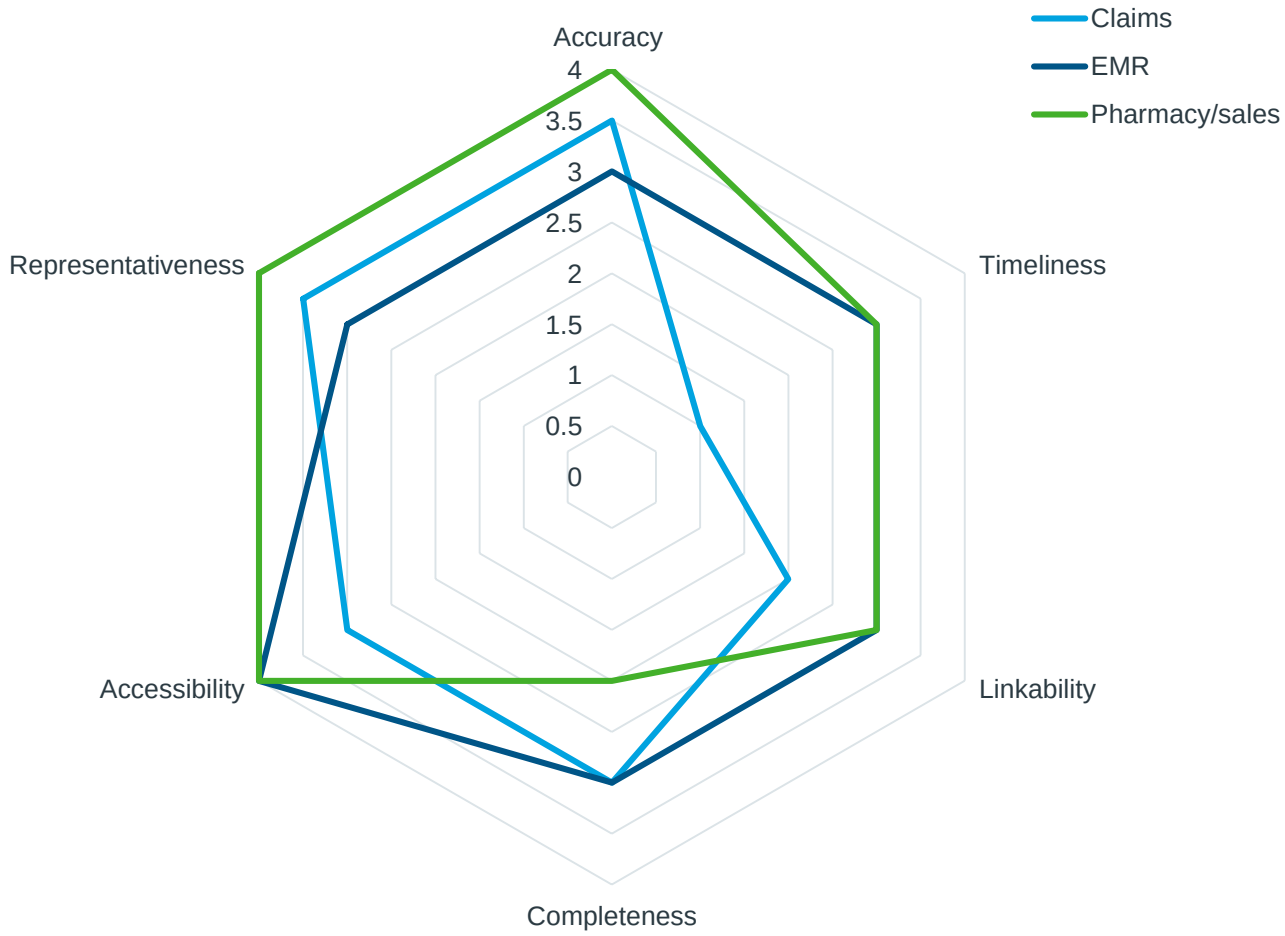
- Demographics
- Drugs (packages sold)

Settings

- Retail pharmacies
- Wholesales
- Company outputs

Good for

- Exposure
- Treatment dynamics (switch, discontinuation…)
- Population movements
- Linkage



Registries

Initial purpose

- Research

Content

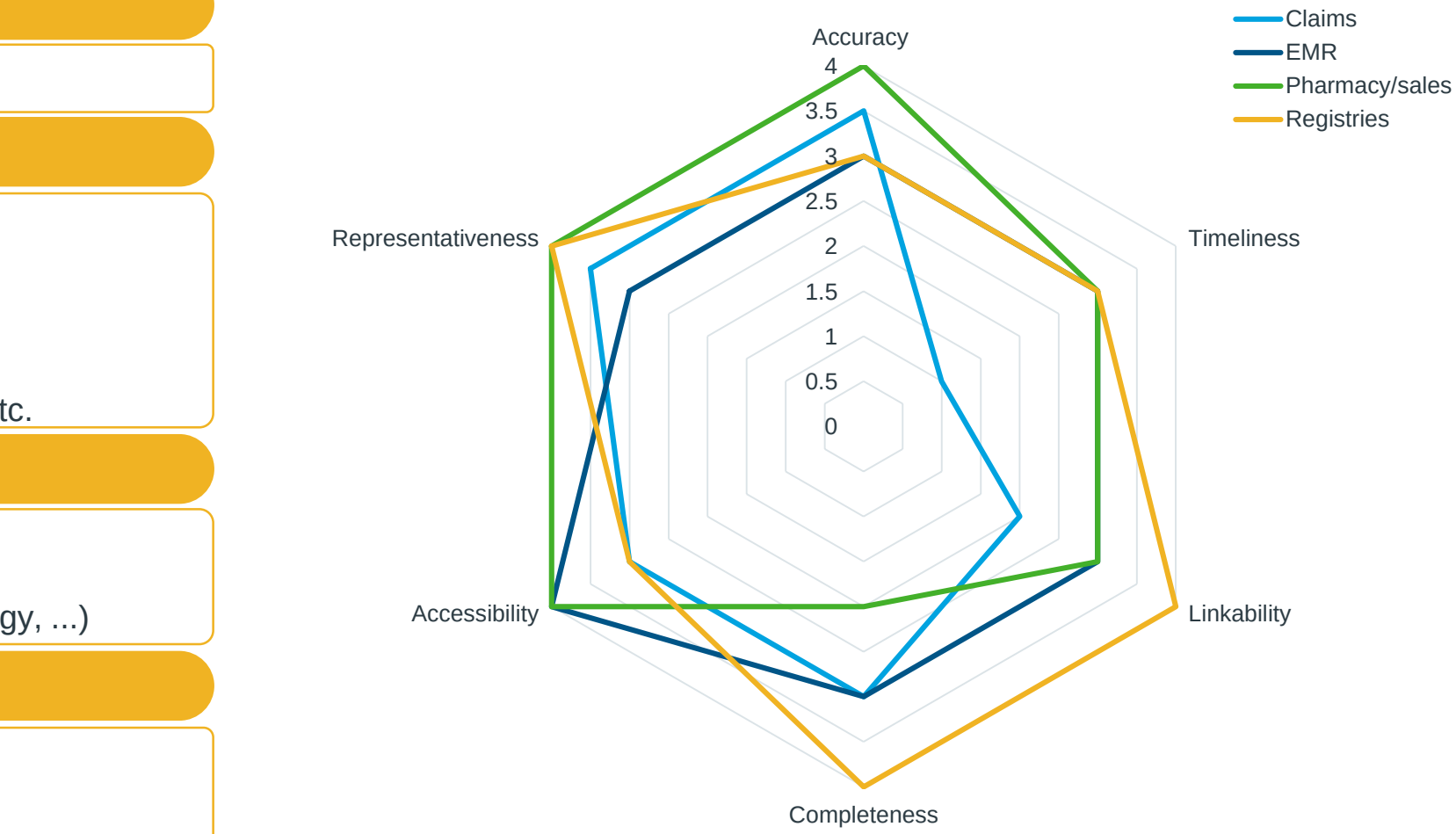
- Demographics
- Clinical details
- Procedures
- Drugs
- Lab values
- Relevant markers, genetic data, tests, etc.

Settings

- Disease or drug oriented
- Mostly secondary care
- Mostly research intensive areas (oncology, ...)

Good for

- Disease epidemiology
- Benefit-risk assessment
- Treatment pathways



Social media, wearables, connected devices, etc

Initial purpose

- Networking
- Follow up
- Experimental

Content

- Demographics
- Narrow and specific data

Settings

- Everyday life
- Smoothly entering the healthcare system
- Telemedicine programs

Good for

- Hypothesis generation
- Signal detection / monitoring
- Population behavior
- Public health intervention evaluation

Data source types dictate only a part of the story

A data landscaping or feasibility assessment is necessary to determine the quality of data

- Databases must be **fit-for-purpose**, i.e. to be able to provide answers to our research questions based on our expectations of access, coverage, validity, likability, timelines, budget, etc.

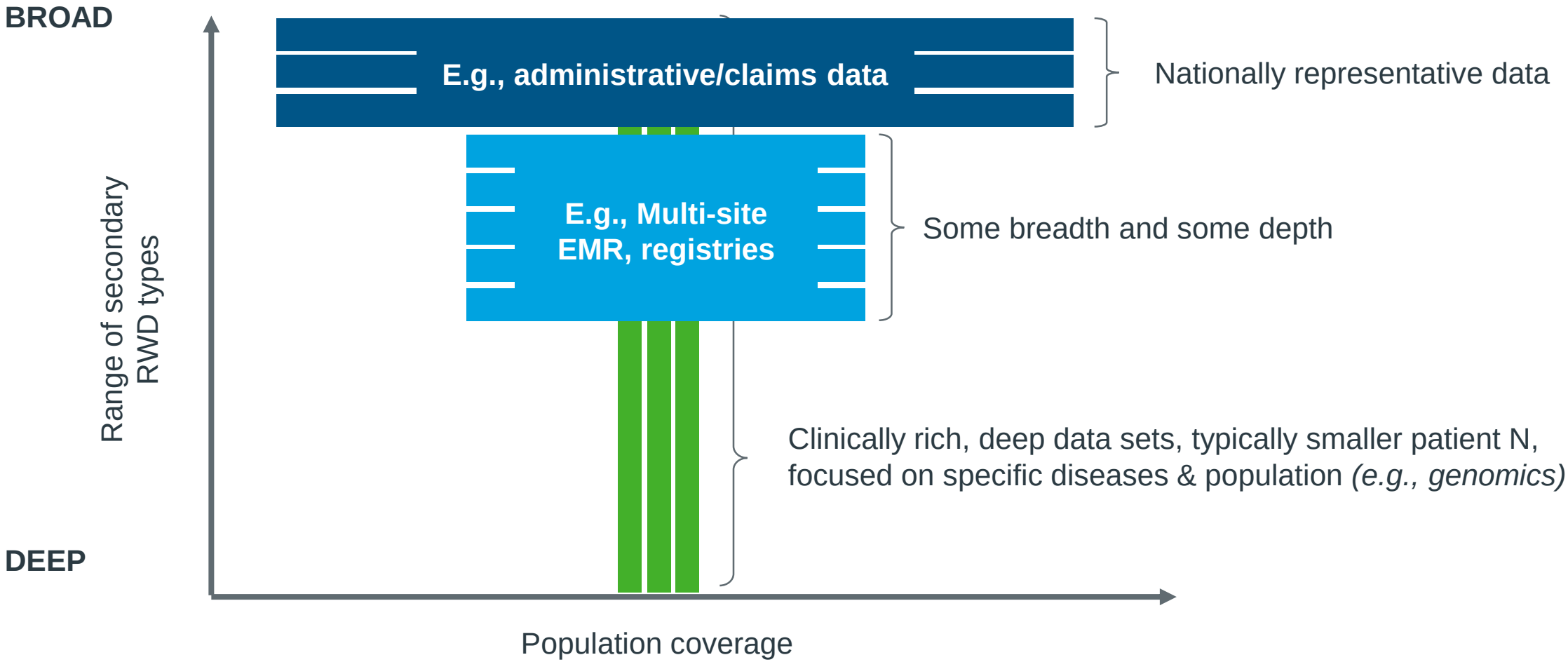
- A **feasibility study** is often necessary to inform the most fit-for-purpose data sources for any given set of research questions.

Criteria	Data Source 1	Data Source 2	Data Source 3	Data Source 4	Data Source 5	Data Source 6	Data Source 7	Data Source 8
	EMR	Database	EMR	Audit	Registry	Biobank	Registry	Registry
Demographics								
Methods of diagnosis								
Treatments				Not reported				
Outcomes				Not reported				
Safety and AEs				Not reported				
PROs								
HCRU								

Country	Spain	UK/Ireland	France	Italy	Germany	Denmark	Norway	Sweden	Finland	International
Number of data sources	10	19	11	13	16	5	3	7	5	13
Geographical Setup										
National	4	10	4	6	12	2	1	3	3	
Regional	5	8	4	6	3	2	2	4	1	
Single-site	1	2	3	1	1	1	0	0	1	
Data Source Type										
EMR/EHR	5	5	4	6	3	3	1	2	3	1
Registry	1	5	4	2	3	2	2	3	1	4
Research/Clinical Database	4	7	1	4	6			1	1	7
Claims/Admin Database/Audit		2	2	1	4			1		1

The fit-for-purpose data may need combining and linkage

T-shaped data linkage allows the extrapolation of high-depth data to larger populations.



Thank You!

Massoud Toussi, MD, PhD, MBA

Global category lead, Evidence from Secondary Data
Real World Evidence Solutions

massoud.toussi@iqvia.com
+33 6 07 96 66 50