

Landscape of data sources for real-world observational studies: focus on the French context

OBJECTIVE

To provide a comprehensive overview of data sources and associated regulatory frameworks available for researchers, study sponsors and regulators in order to investigate the use, safety and effectiveness of therapeutic innovations in France.

METHODS

A structured literature review was conducted to identify existing data sources relevant for real-world observational studies. Both local and international initiatives were examined to ensure the completeness of the review. Given the large number of data sources, we refined the scope of work to France.

RESULTS

The review identified four main categories of data sources which can be re-used for research

1. Primary data collection including data collected for a specific study, by clinical research associates (CRAs) or digital tools such as PRO (patient queries)
2. Reuse of reimbursement data including medico-administrative databases from hospital and primary care reimbursement
3. Reuse of healthcare data collected during patient care or directly collected for or scientific research purposes, including access to hospital or primary care data warehouses.
4. Finally, environmental and socio-economic data, publicly available on national websites or collected through sensors, can be linked to previous data sources using geographic identifiers to enrich analyses. We focused only on open data access as the other data sources are too specific.

Regulatory requirements vary according to the type of data source but most of them involve obtaining authorization from the data owner and approval from ethics committee.

CONCLUSION

This work provides practical guidance for researchers in selecting the most suitable data sources for real world studies and outlines the key regulatory processes.

RESULTS

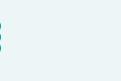
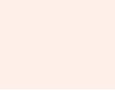
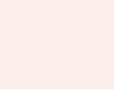
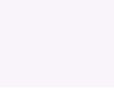
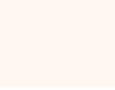
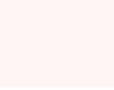
Overview of data sources and associated regulatory frameworks and access terms

Primary data collection		Reuse of data collected for reimbursement		Reuse of data collected in the context of care or research			Environmental and socio-economic data
Primary data collection	Reuse of data collected for reimbursement (hospital)	Reuse of data collected for reimbursement (primary care)	Hospital data	Primary care data	Epidemiological registries	Opendata	
Data categories available							
 Examples in France (not exhaustive)	Adhoc	PMSI, SNDS, DCIR CépiDC	HUGO, APHP, CHU Nantes, CHU Reims...	BIOGROUP	URO CCR, COLIBRI	ATMO, Météo France, CITEPA	
 Access and regulation	CNIL authorization, local ethics and scientific committee	CNIL authorization, protocol, CESREES / HDH validation	CNIL authorization, local ethics and scientific committee	Open access / no regulatory process			
 Time frame	18-24 months	12-18 months	12-18 months	Open access :)			
 Data access costs		No access fees			Open access and no access fees		
 Representativeness	Adhoc	 National	 Hospital / inter-hospital	 According to locations of private laboratories	Adhoc	 National	

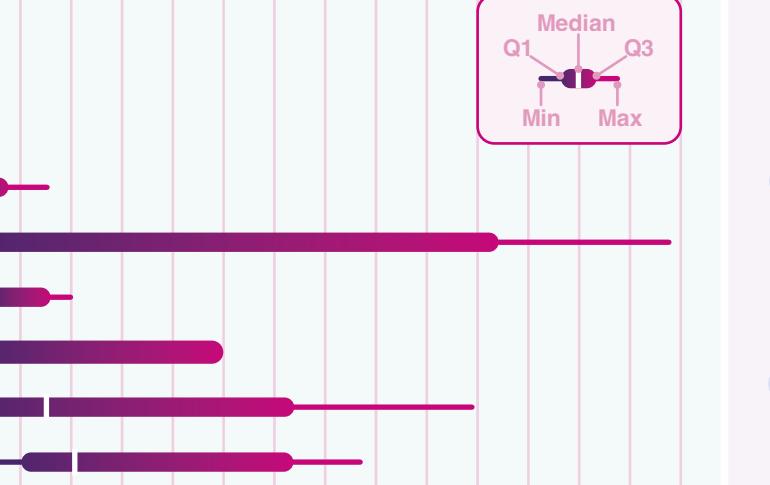
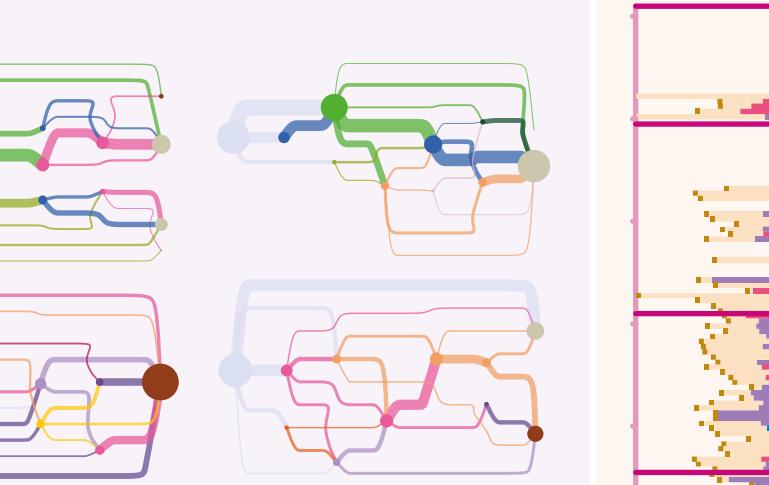
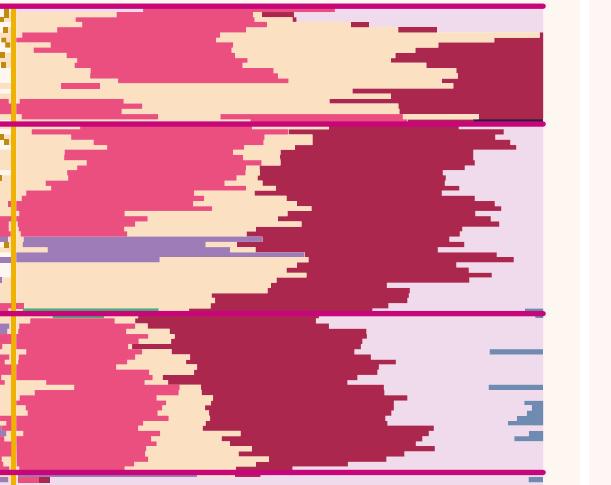
Data categories available for each source and linkage potential

Primary data collection		Reuse of data collected for reimbursement		Reuse of data collected in the context of care or research			Environmental and socio-economic data	
Primary data collection		Reuse of data collected for reimbursement (hospital)	Reuse of data collected for reimbursement (primary care)	Hospital data	Primary care data	Epidemiological registries	Open data	
Data categories available								
 Socio-demographics	✓	✓	✓	✓	✓	✓	✓	✓
 Patient pathways		✓	✓					
 Diagnostic and acts	✓	✓		✓	✓	✓	✓	
 Medical test results	✓			✓	✓	✓	✓	
 Drugs	✓	✓	✓	✓	✓	✓	✓	
 Causes of death		✓	✓					
 Cost of stays		✓	✓					
 PRO, qualitative data from patients	✓	✓			✓	✓		
 Other								✓
 Linkage feasibility	+++			+++	++	+		

Pros & cons of using data sources

Primary data collection		Reuse of data collected for reimbursement		Reuse of data collected in the context of care or research			Environmental and socio-economic data
Primary data collection	Reuse of data collected for reimbursement (hospital)	Reuse of data collected for reimbursement (primary care)	Hospital data	Primary care data	Epidemiological registries	Open data	
Data categories available							
 On demand collected data							
 Confirmed diagnostic							
 Structured data							
 Unstructured data							
 Patient follow up							
 Level of authorities recognition							
Uses cases	<ul style="list-style-type: none"> • Need for data generation • High level expectation of evidence for reimbursement, or drug authorization 			<ul style="list-style-type: none"> • Post drug authorization • Price negotiation 		<ul style="list-style-type: none"> • High level expectation of evidence for reimbursement, or drug authorization 	<ul style="list-style-type: none"> • Epidemiological studies, public health

Use cases examples

Primary data collection	Reimbursement databases	Patients records	Environmental and socio-economic data
<p>Use-cases examples</p> <p>Safety drug study</p> 	<p>Patient pathway study</p> 	<p>Evaluation of efficacy and treatment patterns</p> 	<p>Public study on soil contamination</p> 