

EUROPEAN
MEDICINES
AGENCY

Possible sources of heterogeneity and how this can be assessed

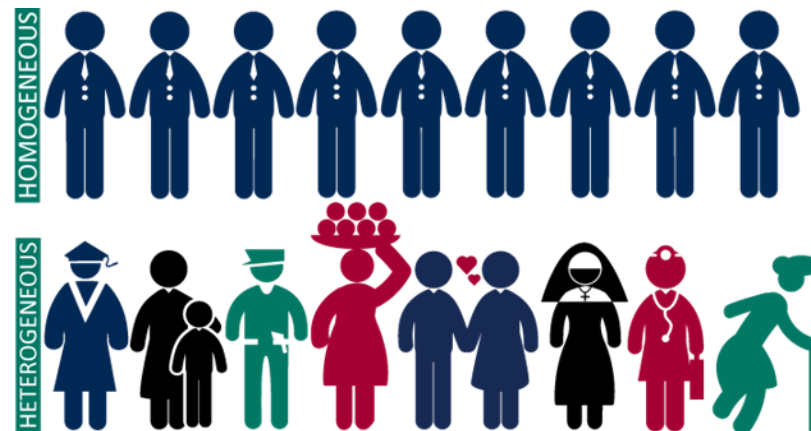
Session: Embracing Diversity and Tackling Heterogeneity in Data, Methods, and Jurisdictions



I have no conflict of interest to declare

1. Sources of heterogeneity in multi database studies (MDS)
2. Analytical methods to account and report heterogeneity
3. EMA use cases
4. Key messages

- Variation in study protocols
- Variation in study quality
- Differences in interventions received (e.g. dose)
- Differences in follow-up length
- Treatment-covariate interaction
- Publication bias
- ...



- A study on prevalence of acute liver injury was requested during a validation meeting for **COMP** to support an **orphan application**
- Prevalence of not more than 5 in 10,000 persons for orphan designation.
- Literature review – rare disease with a prevalence of 1-5 per 10,000 persons.
- What is the yearly prevalence of patients at **risk of acute liver failure**.
- 5 Real World Data Sources:
 - IMS®Disease Analyzer France: Primary care data representing 4,000,000 patients
 - IMS®Disease Analyzer Germany: Primary care data representing 15,000,000 patients
 - THIN Italy: Primary care data representing 900,000
 - THIN Romania: Primary care data representing 500,000
 - THIN Spain: Primary care mainly + hospital data representing 3,000,000

Database	No. of patients with acute liver injury	No. of male patients (%)	Mean (median) age at first acute liver injury diagnosis
IMS®Disease Analyzer France	5785	2903 (50.2%)	51.71 (53)
IMS®Disease Analyzer Germany	244,086	125,950 (51.6%)	52.96 (54)
THIN Italy	3832	2114 (55.2%)	57.61 (58)
THIN Romania	36,411	17,859 (49.0%)	52.33 (55)
THIN Spain	19,086	10,428 (54.6%)	54.70 (55)

Country	Year	Numerator	Denominator	Prevalence per 10,000 (95% CI)
France	2019	1585	817,127	19.40 (18.45 - 20.38)
	2020	1263	757,192	16.68 (15.77 - 17.63)
Germany	2019	67,180	4,617,603	145.49 (144.40 - 146.58)
	2020	58,956	4,303,772	136.99 (135.89 - 138.09)
Italy	2019	1073	583,366	18.39 (17.31 - 19.53)
	2020	1005	596,571	16.85 (15.82 - 17.92)
Romania	2019	10,468	623,169	167.98 (164.80 - 171.20)
	2020	8783	584,436	150.28 (147.18 - 153.43)
Spain	2019	5102	1,152,732	44.26 (43.06 - 45.49)
	2020	4091	1,118,529	36.57 (35.46 - 37.71)

- Different coding systems
 - ICD 10 (Germany, France, Romania) vs ICD 9 (Italy and Spain)
- Different national health surveillance systems (i.e. screening and testing practices) and (health seeking) behaviour
 - Chronic hepatitis C (23,5%) in Italy and Acute hepatitis C (17,1%) in Spain
 - Prevalence of Hepatitis C is 0.1% to 5.9% in Europe
- Different quality of recording
 - Recording of diagnosis less accurate in primary care setting in France and Italy
 - ICD 10 = K76.0 Fatty (change of) liver not elsewhere classified
- Different organisation of health care (primary versus secondary)

1) Methodological heterogeneity

- Measurement heterogeneity (i.e. clinical classification system)
- Information heterogeneity (i.e. granularity of clinical codes)

➡ unwanted heterogeneity

2) Clinical heterogeneity ➡ informative

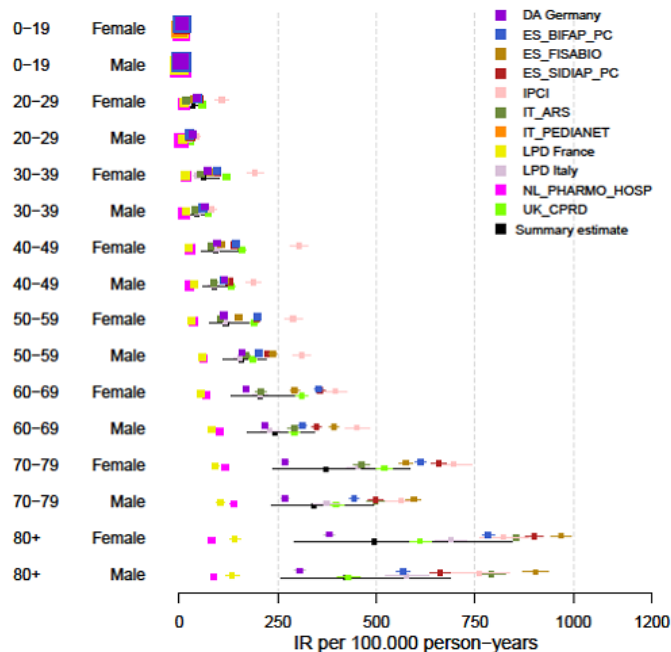
- Improving external validity
- Differences in prescription patterns between different geographical regions, subpopulations or behaviours
- Impact of risk minimisation measures (RMMs) in different health care systems

- COVID-19 pandemic, RWE provided evidence on safety of COVID-19 vaccines
- Background incidence rates of adverse events of special interest (AESI) in 13 databases
- VTE (venous thromboembolism) as example – Vaxzevria
 - **Clinical heterogeneity**
 - By gender and prespecified age groups
 - **methodological heterogeneity**
 - ACCESS with 12, ERASMUS with 6 (2 consortia, ACCESS & ERASMUS)
 - Different studies designs/study population.
 - Different phenotyping.
 - RWD data characteristics
 - Some PC only, some PC with hospital linkage, one possibly hospital only.
 - Different classification system (ICD10, SNOMED)

- Visualising heterogeneity using forest plots
- Quantifying heterogeneity using statistical methods
 - Frequentist random effects model
 - Summary estimate
 - ❑ τ^2 is an estimate of the variance of the true effect sizes
 - ❑ I^2 is a measure for the proportion of observed variance that reflects real differences in effect size.
- Supplementary analyses
 - ❑ Only those with same case definition (ERASMUS vs. ACCESS)
 - ❑ Only those linkage between primary care and hospital data
 - ❑ Only those with same clinical classification system (ICD-10)

Results: Summary estimate

Age-gender-stratified IR estimates and CI for VTE by database and pooled



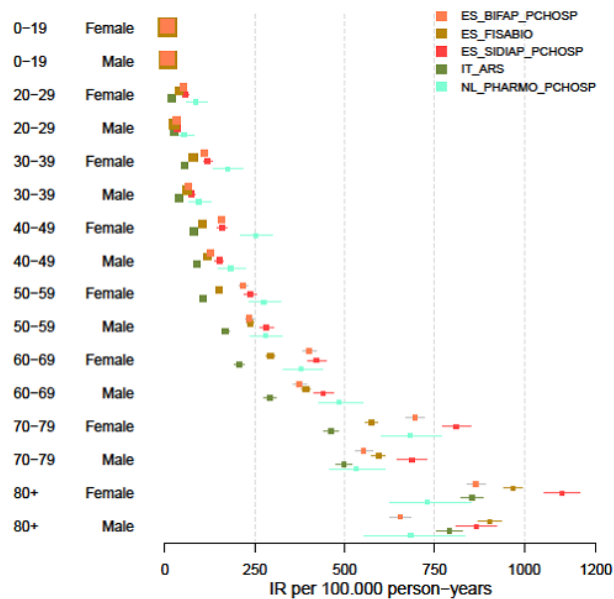
**Estimated IR of VTE from 5 to 421 per
100,000 person-years from meta-analysis**

I^2 and τ^2 from meta-analyses

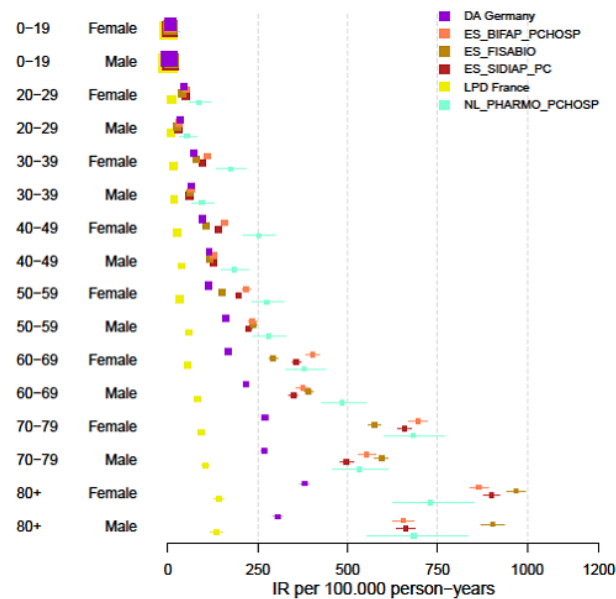
Age	Gender	I^2	τ^2
0-19	Female	0.872	0.253
0-19	Male	0.921	0.376
20-29	Female	0.983	0.531
20-29	Male	0.954	0.381
30-39	Female	0.992	0.626
30-39	Male	0.981	0.423
40-49	Female	0.996	0.600
40-49	Male	0.991	0.367
50-59	Female	0.997	0.517
50-59	Male	0.995	0.299
60-69	Female	0.998	0.473
60-69	Male	0.996	0.309
70-79	Female	0.998	0.527
70-79	Male	0.997	0.358
80+	Female	0.998	0.735
80+	Male	0.997	0.625

Age-gender-stratified IR estimates and CI for VTE by database

With PC-H linkage



ICD-10 classification



1. Using a multi-database approach provides a **more accurate picture** of true incidence rates, as there may be large clinical differences underpinning the variability in the estimates across different databases.
2. After **mitigating unwanted heterogeneity** through harmonization of database characteristics, there might still be some heterogeneity present, but this should be considered as a source of knowledge;
3. When heterogeneity is present, a careful trade-off has to be made between **stratified estimates or a pooled estimate**, to support use in pharmacoepidemiological and regulatory evaluation.
4. To attenuate heterogeneity, a pre-screening of database characteristics through **a meta-dataset** and adequate analytical tools at study design stage might be considered.

Any questions?

Further information

chantal.quinten@ema.europa.eu

Official address Domenico Scarlattilaan 6 • 1083 HS Amsterdam • The Netherlands

Telephone +31 (0)88 781 6000

Send us a question Go to www.ema.europa.eu/contact

Follow us on  **@EMA_News**

TABLE 1 Overview of main characteristics by data source.

	Consortium	Region	Type of data source/health care system: Primary care (PC) or Hospital (H) data	Study population	Clinical Classification Coding system
PHARMO	ACCESS	Netherlands	H	9 184 832	ICD10
PHARMO	ACCESS	Netherlands	PC and H	496 197	ICPC (PC), ICD10 (H)
BIFAP	ACCESS	Spain	PC	10 266 468	SNOMED, ICD9
BIFAP	ACCESS	Spain	PC and H	4 423 843	SNOMED, ICD9 (PC), ICD10 (H)
SIDIAP	ACCESS	Catalonia (Spain)	PC	6 205 573	ICD10
SIDIAP	ACCESS	Catalonia (Spain)	PC and H	1 758 239	ICD10
FISABIO	ACCESS	Valencia region (Spain)	PC and H	5 886 560	ICD9, ICD10
PEDIANET	ACCESS	Italy	PC	181 290	ICD9
ARS	ACCESS	Tuscany (Italy)	PC and H	3 067 602	ICD9
CPRD GOLD	ACCESS	United Kingdom	PC	4 688 710	READ, SNOMED
CPRD GOLD	ERASMUS	United Kingdom	PC	3 913 071	READ
IQVIA DA Germany	ERASMUS	Germany	PC	8 459 098	ICD10
IQVIA LPD France	ERASMUS	France	PC	3 951 633	ICD10
IPCI	ERASMUS	Netherlands	PC	1 299 288	ICPC
IQVIA LPD Italy	ERASMUS	Italy	PC	1 066 230	ICD9
SIDIAP	ERASMUS	Catalonia (Spain)	PC and H	1 909 814	ICD10

Abbreviations: ARS, Agenzia Regionale di Sanita Toscana; BIFAP, Base de Datos para la Investigacion Farmacoepidemiologica en Atencion Primaria; CPRD, clinical practice research datalink; DA, disease analyzer; FISABIO, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana; IPCI, integrated primary care function; LPD, longitudinal patient data; SIDIAP, Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària.

With thanks to the colleagues for providing the use cases:

- Karin Hedenmalm
- Juan Jose Andres Abellan
- Xavier Kurz
- Martin Russek

References:

- [Data Analysis and Real World Interrogation Network \(DARWIN EU\) | European Medicines Agency \(europa.eu\)](https://europeanmedicinesagency.europa.eu/data-analysis-and-real-world-interrogation-network-darwin-eu)
- Russek M, Quinten C, de Jong VMT, Cohet C, Kurz X. Assessing heterogeneity of electronic health-care databases: A case study of background incidence rates of venous thromboembolism. *Pharmacoepidemiol Drug Saf.* 2023 Sep;32(9):1032-1048. doi: 10.1002/pds.5631. Epub 2023 May 8. PMID: 37068170