European Medical Information Framework

Symposium

Friday 22nd September 2017

Instituto de Investigación
Hospital 12 de Octubre

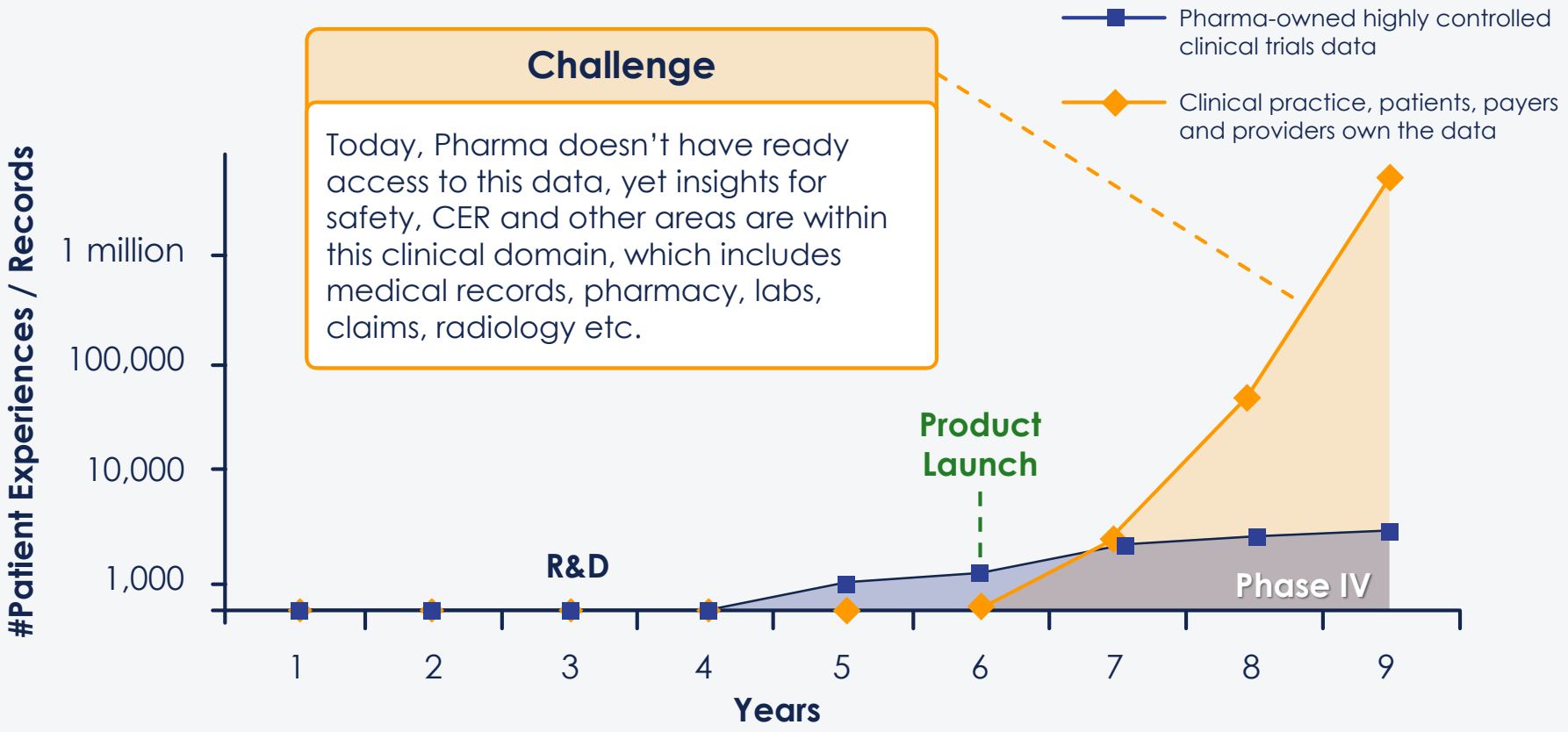
Introductory Welcome

Bart Vannieuwenhuyse

Janssen Pharma R&D

Why is EMIF needed?
Secondary use of health data to enrich research

The "burning platform" for life sciences



The value of healthcare data for secondary uses in clinical research and development — Gary K. Mallow, Merck, HIMSS 2012

Project overview

ACADEMIC PARTNERS 37

Universitätsklinikum Erlangen · University of Helsinki · Brighton · Erasmus MC · EuroRec · EMBL European Molecular Biology Laboratory · ARS Toscana · University of Gothenburg · King's College London · University of Copenhagen · Assistance Publique Hôpitaux de Paris · University of Oxford · Itä-Suomen yliopisto · Universiteit Antwerpen · Universitat Pompeu Fabra Barcelona · Vestische Kinder- und Jugendklinik Datteln · Manchester 1824 · Karolinska Institutet · Inserm · VTT · University of Leicester · University of Cambridge International Examinations · Università di Pisa · VIB · midt Aarhus Universitetshospital · VUmc · UCL · University of Glasgow · IDIAP Jordi Gol · UPMC

SME PARTNERS 9

cambridge cognition · concentris · genomedics · gnúbila · PHARMO · CUSTODIX · proteome sciences · SYNAPSE · pedianet

EFPIA PARTNERS 10

AMGEN · janssen · Roche · ucb · novo nordisk · MERCK · Pfizer · Boehringer Ingelheim · SERVIER · gsk GlaxoSmithKline

PATIENT ORGANISATION 1

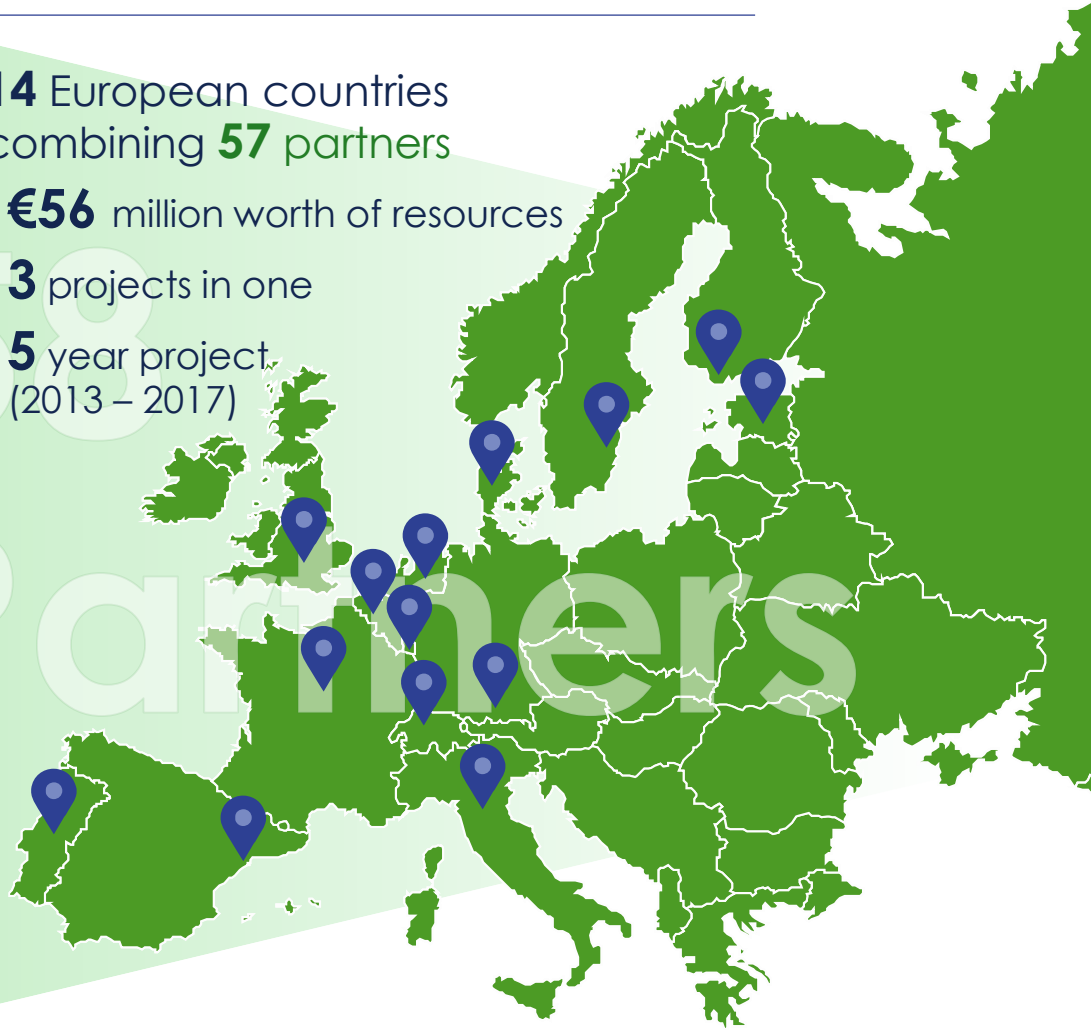
Alzheimer Europe

14 European countries combining **57** partners

€56 million worth of resources

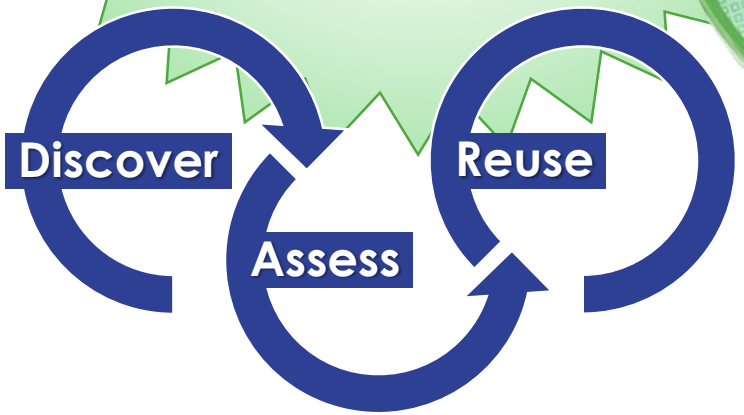
3 projects in one

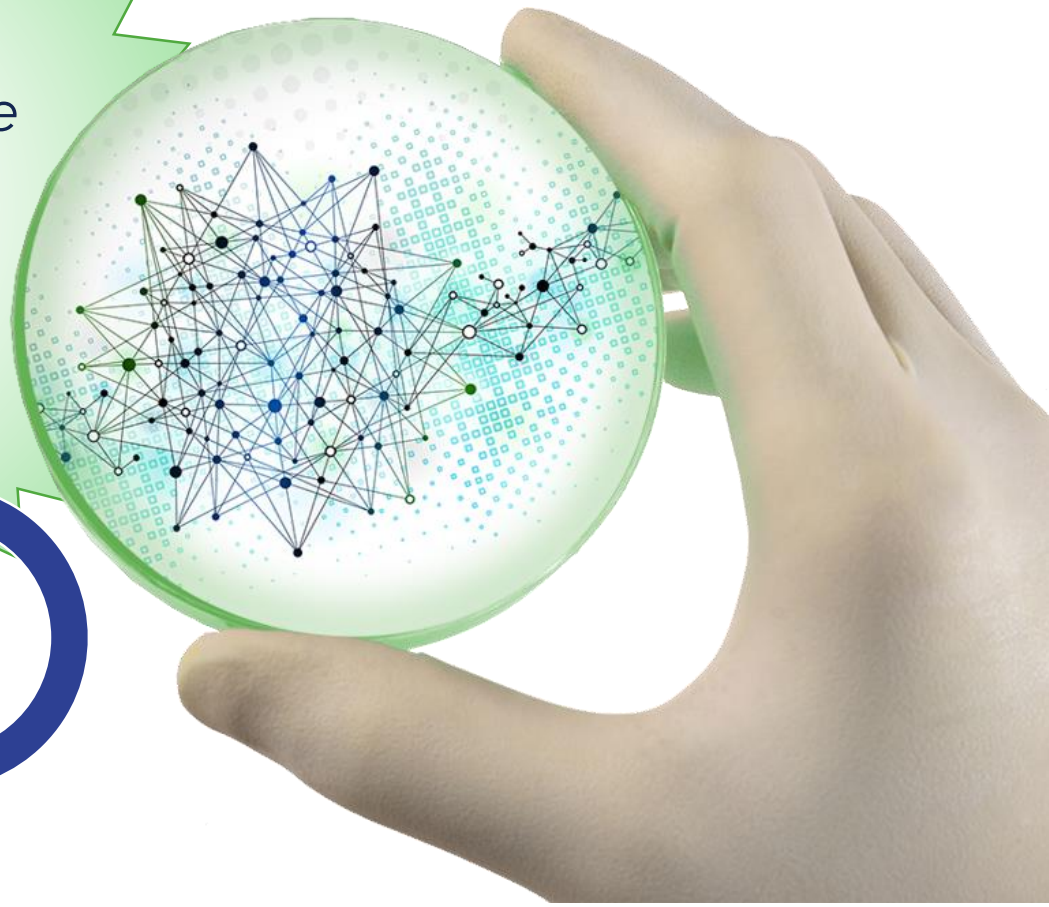
5 year project (2013 – 2017)



Our vision

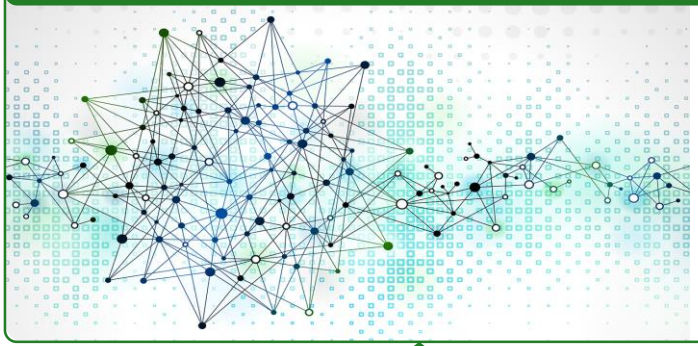
To become the trusted European hub for health care data intelligence, enabling new insights into diseases and treatments





Project objectives

EMIF-Platform

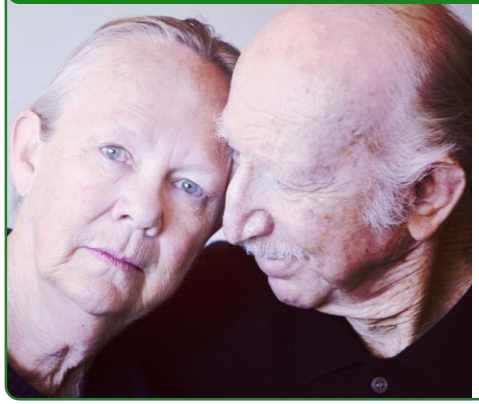


Develop a framework for evaluating, enhancing and providing access to human health data across Europe, support EMIF-Metabolic and EMIF-AD (the specific topics below) as well as support research using human health data in general

EMIF-Metabolic

Identify predictors of metabolic complications in obesity

EMIF-AD



Identify predictors of Alzheimer's Disease (AD) in the pre-clinical and prodromal phase

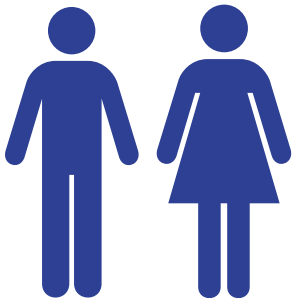
Available data types

Large variety in "types" of data

- Primary care data sets
- Hospital data
- Administrative data
- Regional record-linkage systems
- Registries and cohorts (broad and disease specific)
- Biobanks
- Secondary care data sets
- Paediatric data sets

Data is available from more than 40 million subjects from six EU countries, and in addition:

25,000 subjects in AD cohorts

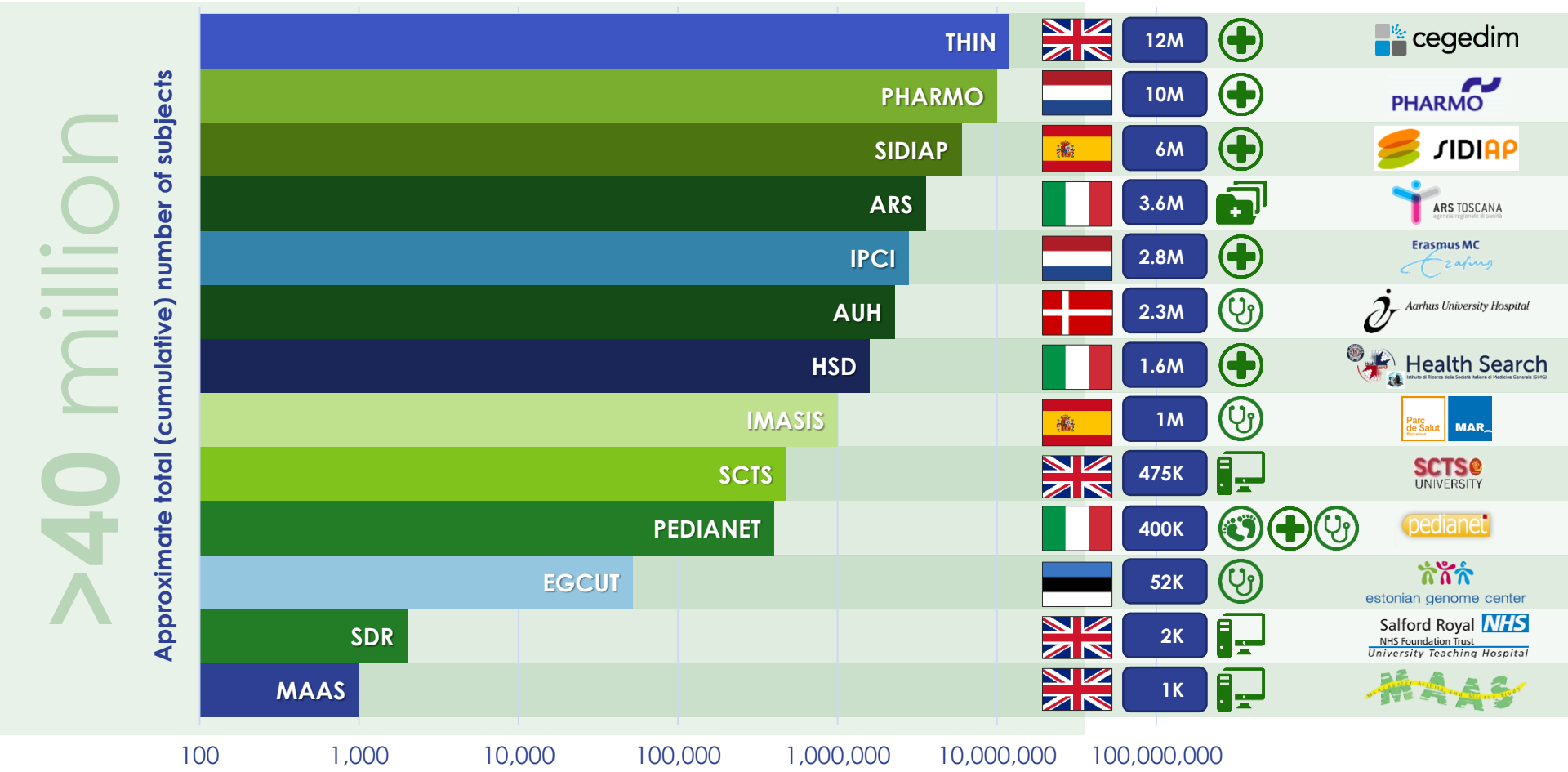


more than **94,000** subjects in metabolic cohorts

Available data sources

EMIF-Available Data Sources; *EXAMPLES*

Status Jan 2016

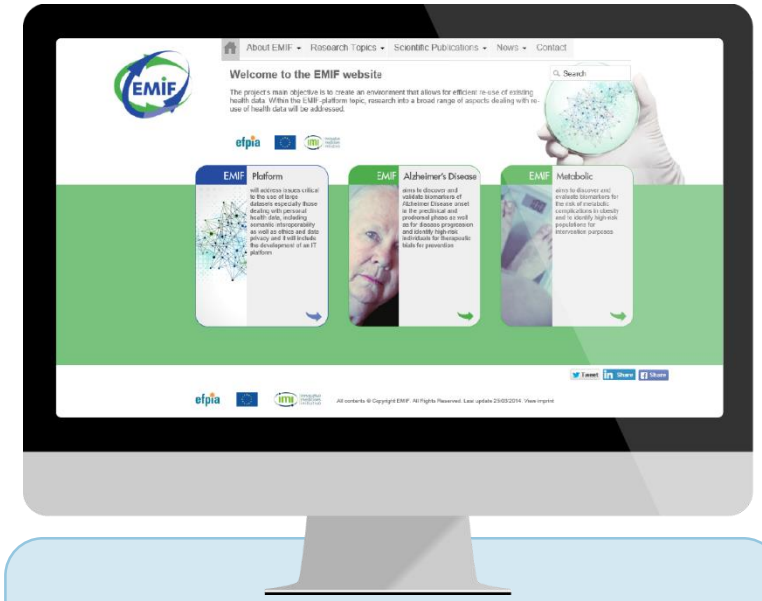


Today at a glance --

- Results from EMIF work to date
- Projecting forward
- Power of data harmonization – OHDSI experience
- Bring it all together – panel discussion

More Information

- **EMIF general**
  - Bart Vannieuwenhuyse (bvannieu@its.jnj.com)
  - Simon Lovestone (simon.lovestone@psych.ox.ac.uk)
  - Johan van der Lei (j.vanderlei@erasmusmc.nl)
- **EMIF-Platform**
  - Johan van der Lei (j.vanderlei@erasmusmc.nl)
  - Nigel Hughes (nhughes@its.jnj.com)
- **EMIF-Metabolic**
  - Ulf Smith (ulf.smith@medic.gu.se)
  - Dawn Waterworth (dawn.m.waterworth@gsk.com)
- **EMIF-AD**
  - Pieter Jelle Visser (pj.visser@maastrichtuniversity.nl)
  - Johannes Streffer (jstreffe@its.jnj.com)



www.emif.eu
Register for our newsletter!

EMIF is operating under IMI Grant Agreement n°115372

Research Use Cases – What Have We Learned?

Chair: Prof Simon Lovestone

Oxford University

EMIF Metabolic

Bart Vannieuwenhuyse

Janssen Pharma R&D

EMIF-Metabolic

Bart Vannieuwenhuyse
(on behalf of the EMIF-Metabolic team)

September 2017

Project objectives – EMIF-metabolic

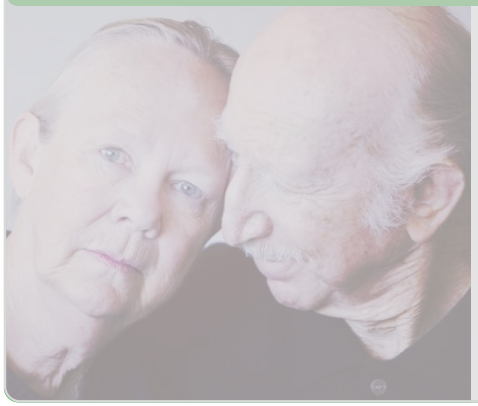
EMIF-Platform

Develop a framework for evaluating, enhancing and providing access to human health data across Europe, support EMIF-Metabolic and EMIF-AD (the specific topics below) as well as support research using human health data in general

EMIF-Metabolic

Identify predictors of metabolic complications in obesity

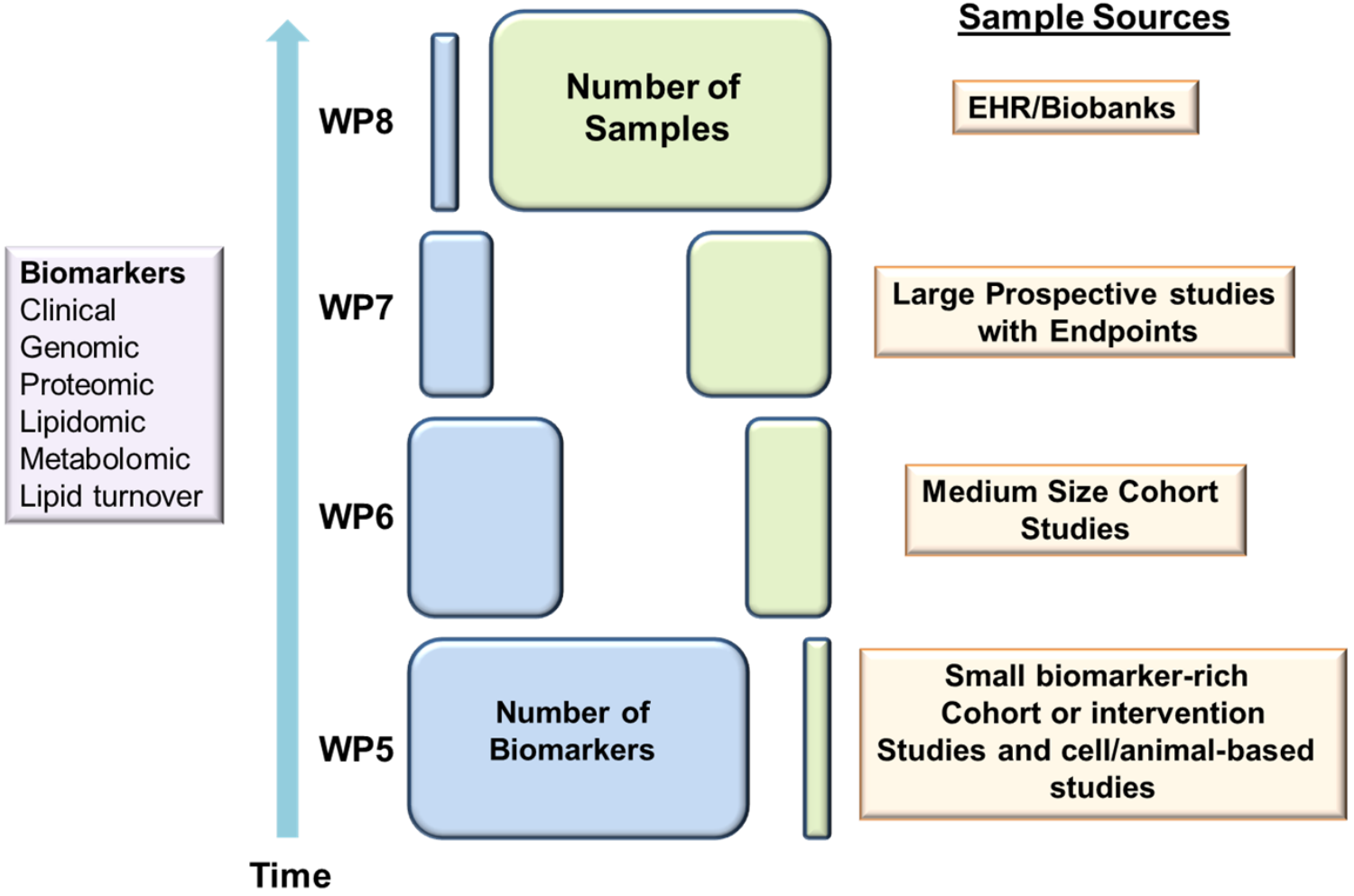
EMIF-AD



Identify predictors of Alzheimer's Disease (AD) in the pre-clinical and prodromal phase

efpia imi innovative medicines initiative

EMIF-Metabolic: objectives

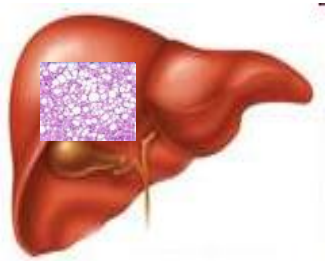


WHAT IS NAFLD ?

(non-alcoholic fatty liver disease)

- Ectopic liver fat from excess consumption of calories arising when safer fat stores are over-filled

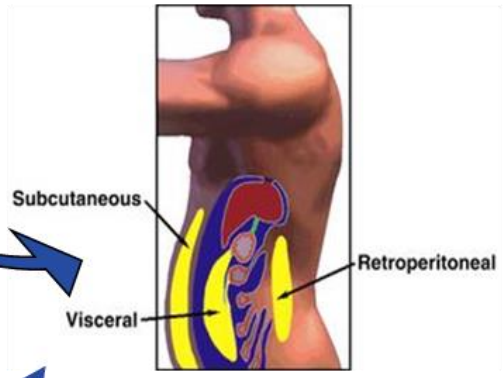
- NAFLD also risk factor for more severe liver complications

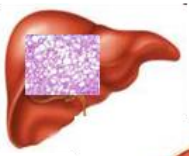


Excess calories
(increased intake or
reduced energy expenditure)

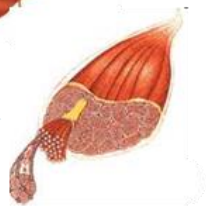
Subcutaneous stores
overwhelmed
(genes / sex, FHx,
ethnicity, ageing)

FAT
'Spill over'

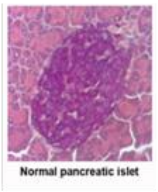




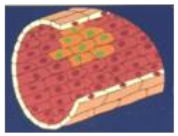
Hepatic lipid
accumulation



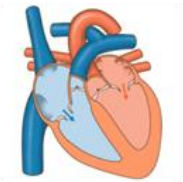
muscle



pancreatic
beta cell



Perivascular fat ⇒
Endothelial dysfunction

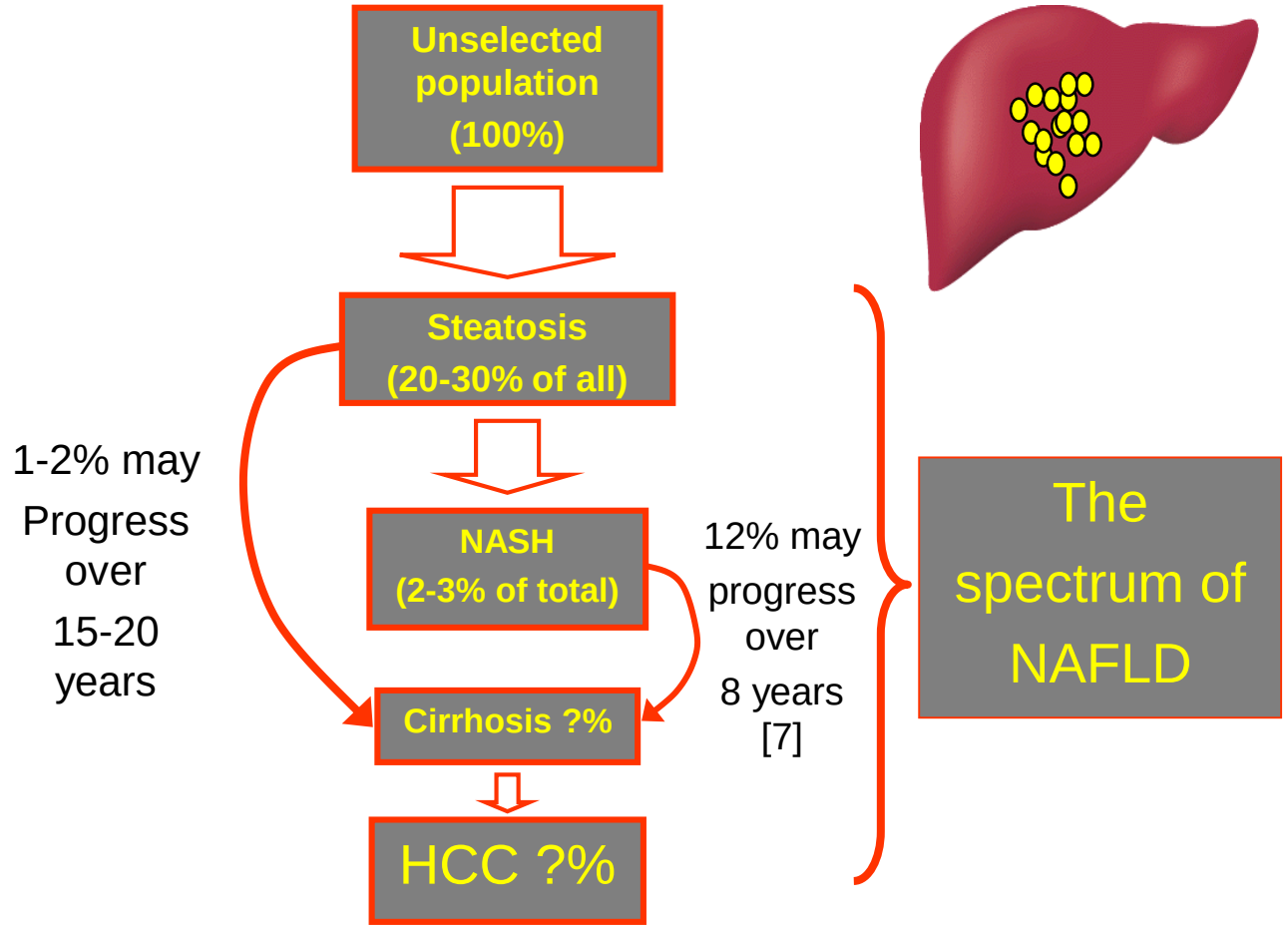


Insulin resistance

Hyperglycaemia

Sattar and Gill
(2014) BMC
Medicine

NAFLD: risk factor for serious disease



Preiss & Sattar Clinical Science 2008

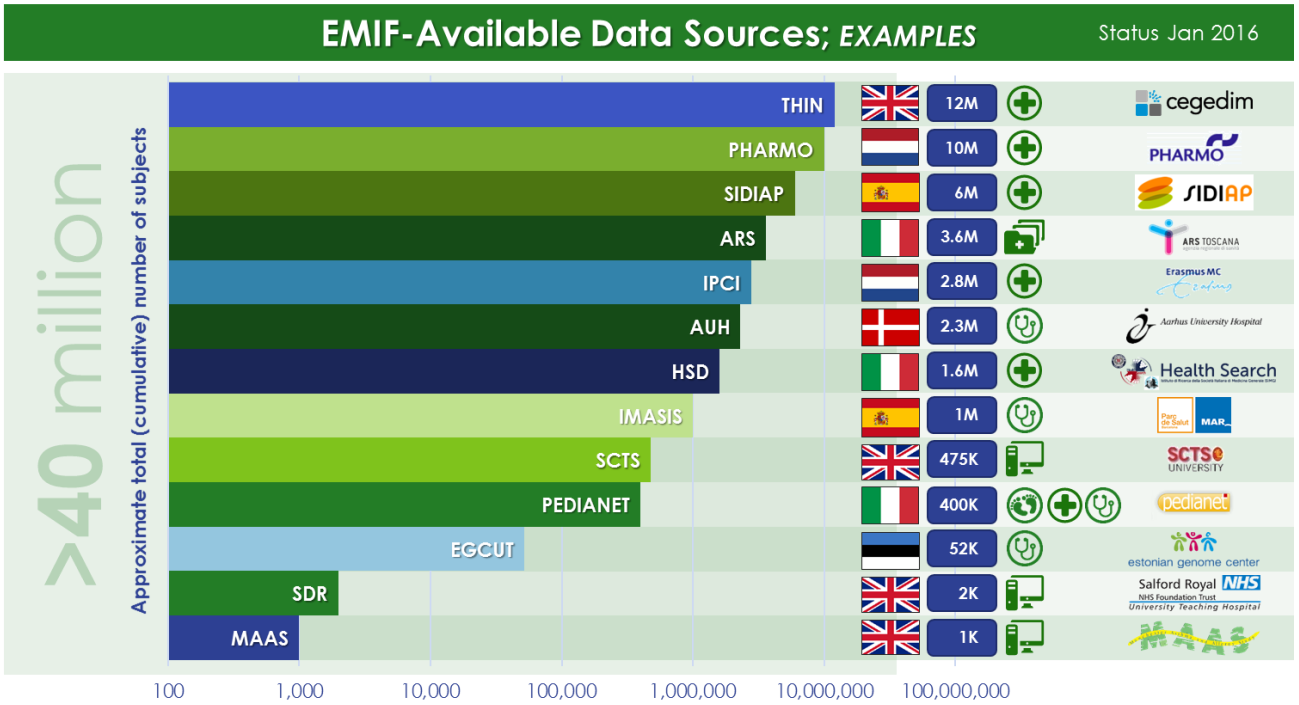
Some examples of big questions being asked

- What is the prevalence of documented non-alcoholic fatty liver disease (NAFLD) disease in clinical practice?
  - Does it vary by country?
  - Is it rising over time?

- Is NAFLD a (strong) risk factor for heart disease?

EMIF- Metabolic – use case

- Use of EHR data – answering the questions
- Findings, Learnings and limitations



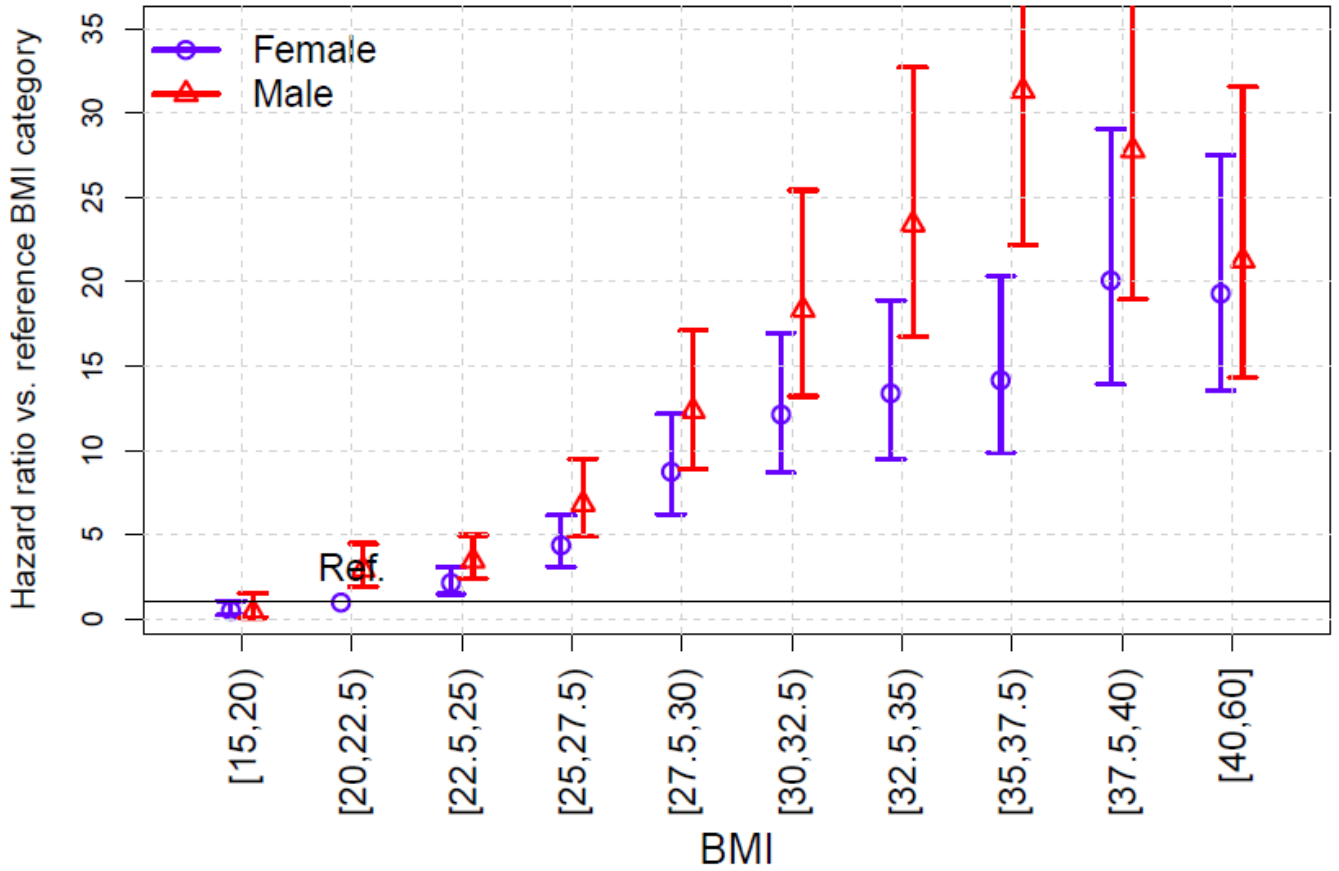
Data outputs – 3 quick examples

- #1 Loomis et al – Risk of NAFLD by baseline BMI in major US / UK datasets –
  - Higher with rising BMI, in diabetes and potentially in men

- #2 prevalence of NAFLD in 4 major EU EHR-datasets
  - Much lower than expected, likely due to under-diagnosis, but prevalence rising

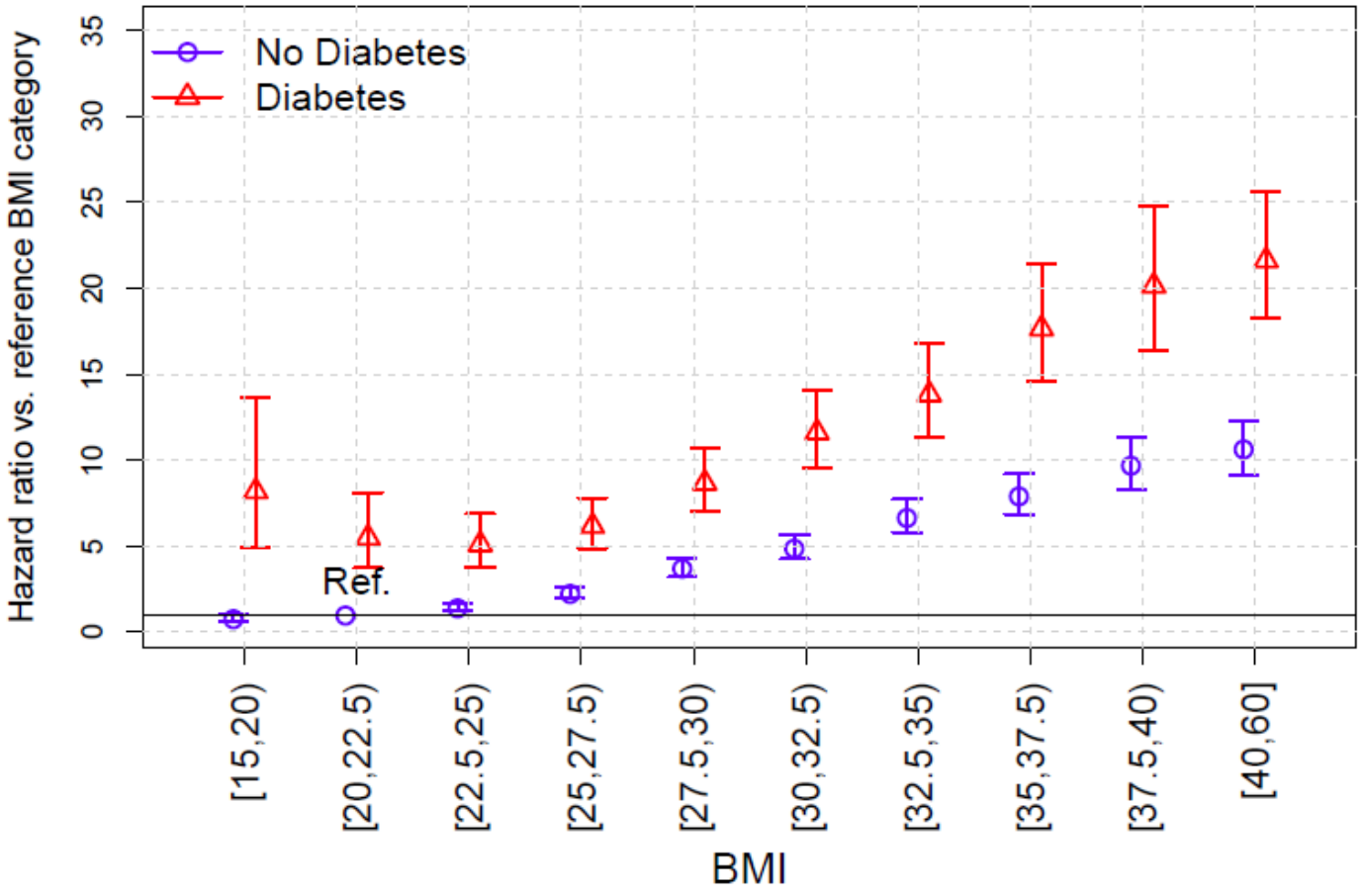
- #3 NAFLD is weak predictor of CVD
  unlikely to be clinically meaningful – goes against some major editorials papers

Hazard ratio NAFLD vs BMI / gender



#1 Loomis, Waterworth, Sattar (2016) JCEM

Hazard ratio NAFLD vs T2DM / BMI

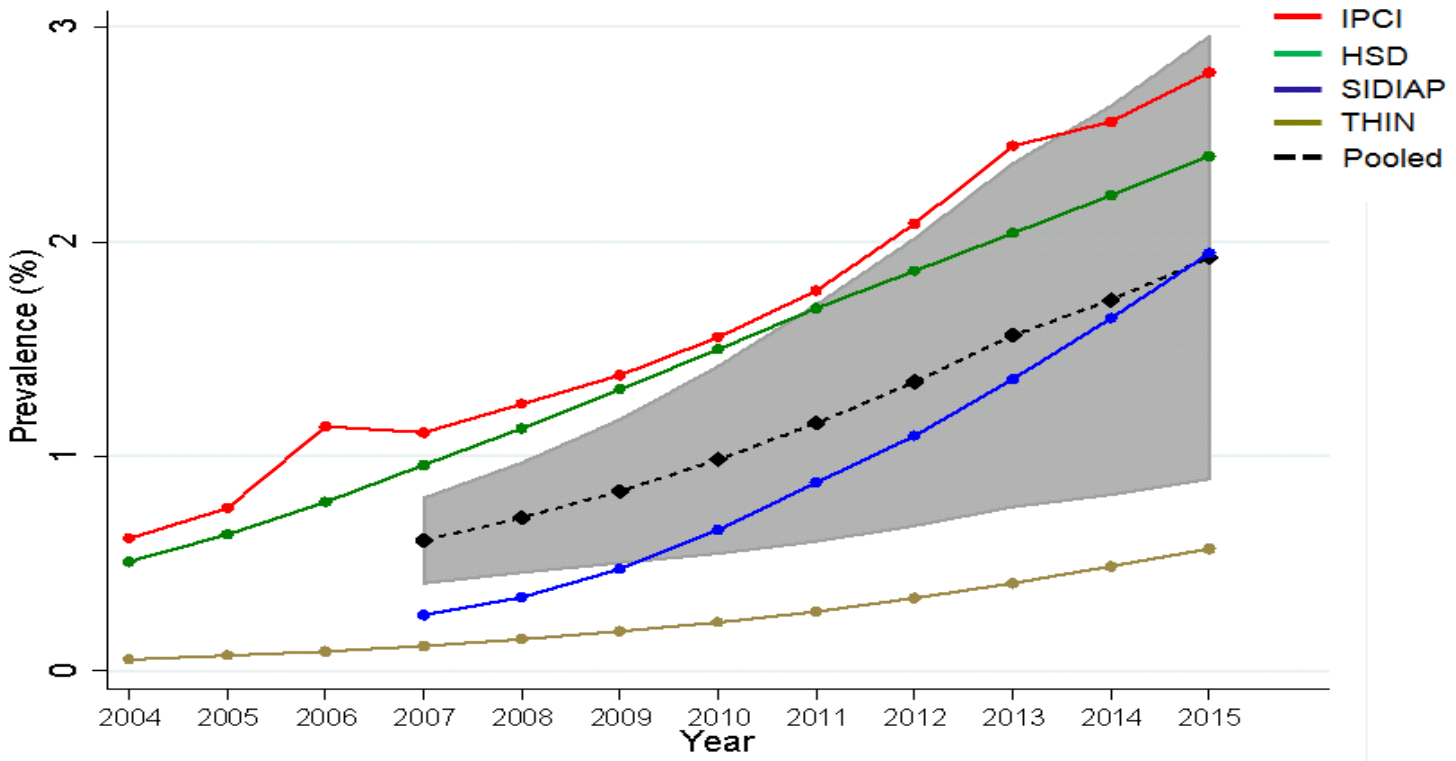


#1 Loomis, Waterworth, Sattar (2016) JCEM

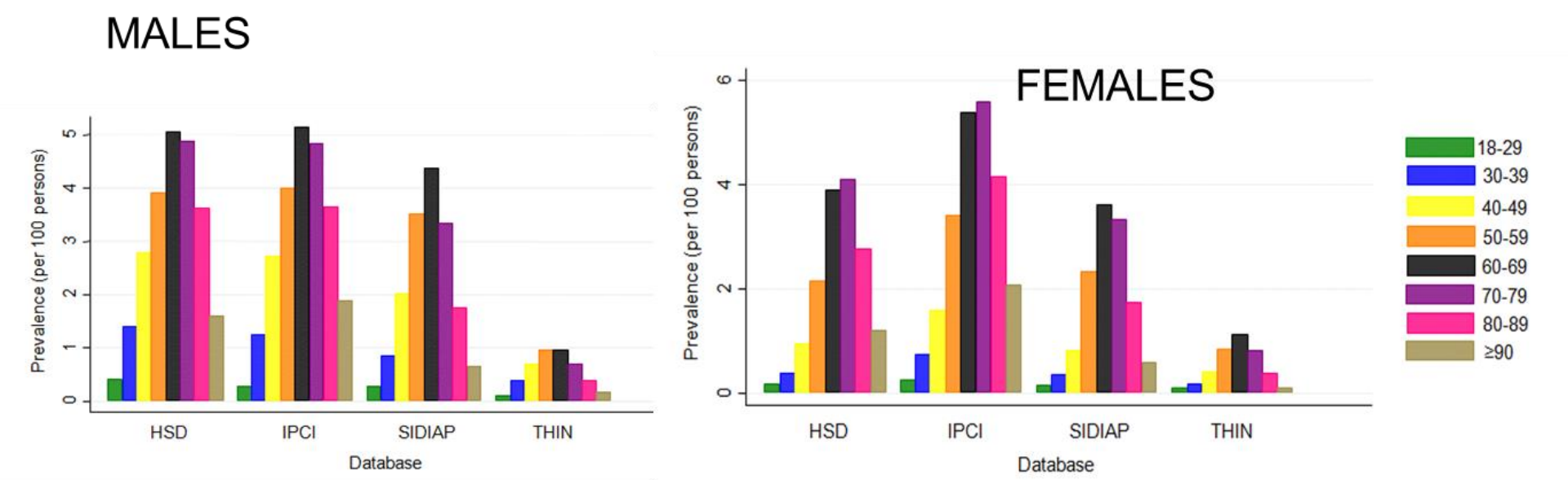
#2 –EMIF – NAFLD prevalence in 4 major EU EHR-datasets – (work near completion)

- Worthwhile question – yes
- Data available – yes
- Collaboration with all relevant parties – yes

Prevalence of NAFLD/NASH



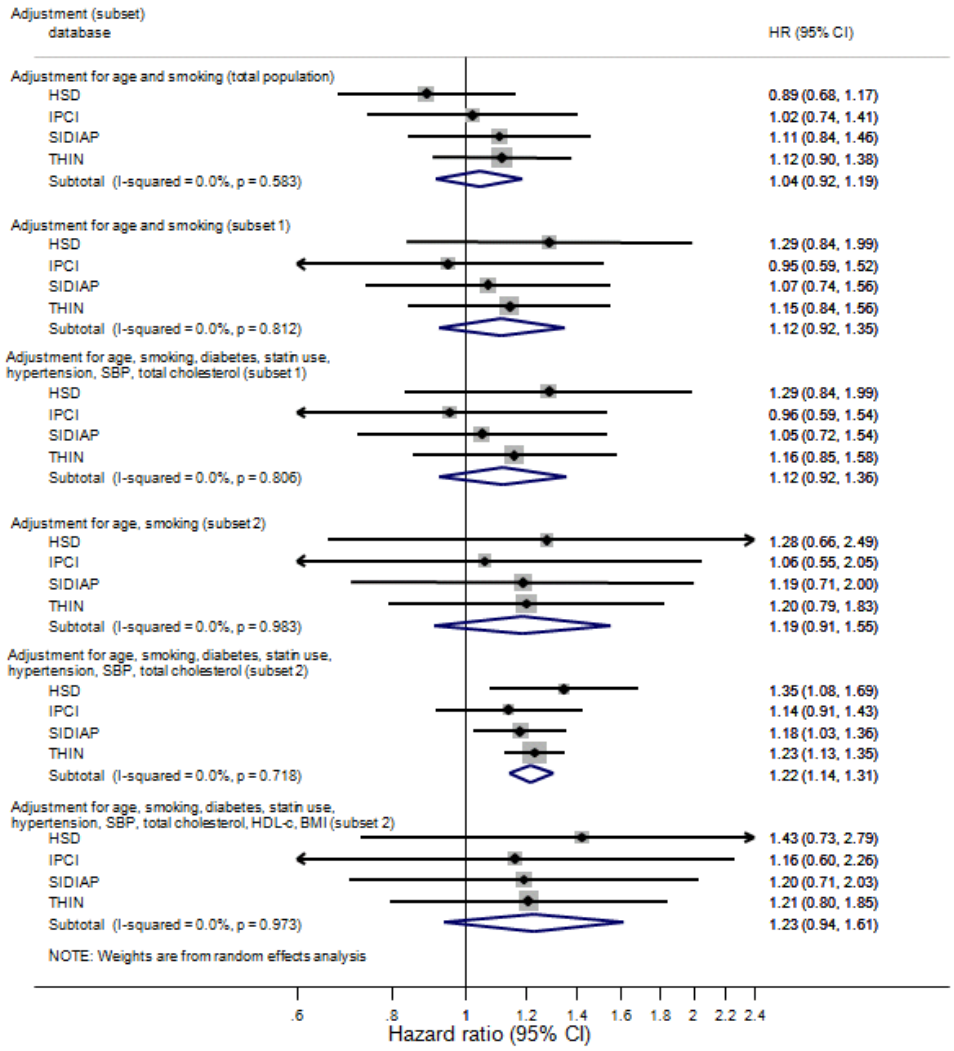
NAFLD Prevalence by gender (on the 1st of Jan 2015)



Higher in men in most datasets and ages – men at higher risk for given BMI

#3 NAFLD & incident MI

- Overall associations modest
- Note findings broadly consistent from 4 major EHRs
- Not able to adjust for more risk factors
- Results important for clinical practice
- NAFLD much stronger risk factor for Diabetes than MI



Experiences gained

- Need to work out importance of question first –
- Can it be delivered from EHR?
  - Do we have right data / sufficient capture of confounders?
  - Do we have robust assessment of outcomes of interest?
  - Do we have sufficient power?
- Works best when data providers, statisticians, scientists / clinicians with relevant epi experience collaborate (need to do this better)
  - Ultra-careful to assure question can be answered with degree of robustness before time and effort expended
  - And, make sure to ask will the answer really take us further?

Limitations

- Often missing data of importance
- E.g BMI commonly missing or measured only on those with risk factors or disease – potential major biases
- Can be overcome but need to be aware
- Reverse causality so longer follow-ups help
- Coding and understanding or outcome measures can be difficult / vary by EHR
- Easy to make simple mistakes, come to potentially wrong / non-robust conclusions

Conclusions

- Takes time to come to grips with EHR derived data
- Many groups need to come together to make important leaps
- Requires time and experience of epi / and understand strengths and limitations of EHRs to make real gains
- Lots of richness but need /experience time to realise


EMIF


efpia
imi
innovative
medicines
initiative

EMIF Alzheimer's

Pieter Jelle Visser

VU Medical Centre, Maastricht

Data sharing in clinical research: the EMIF-AD experience

Pieter Jelle Visser, MD, PhD
Maastricht University
VU University Medical center Amsterdam
The Netherlands

efpia

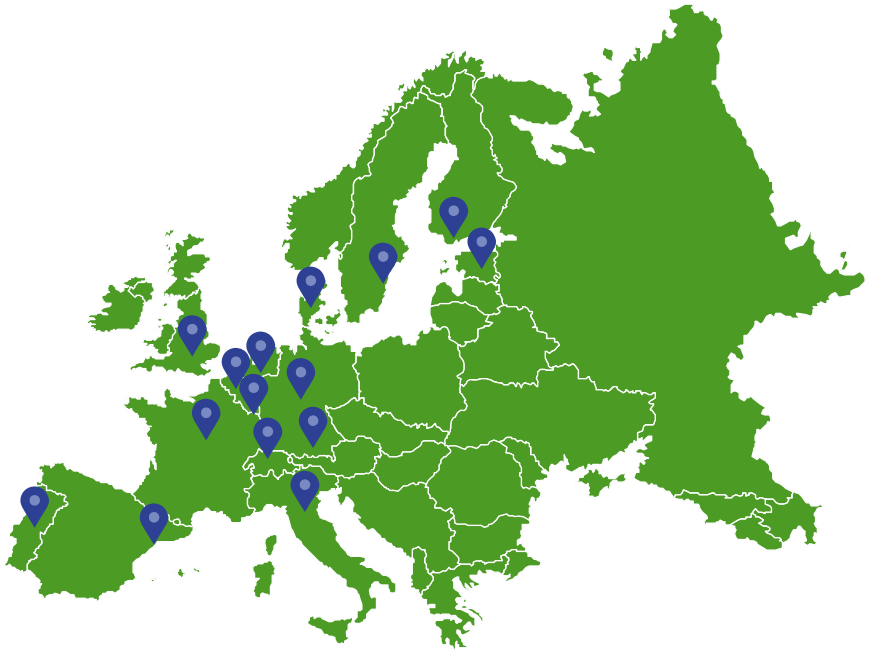
innovative medicines initiative

EMIF

- Vision:
  - European hub for health care data intelligence, enabling new insights into diseases and treatments

- Three subprojects:
  - EMIF-Platform
  - EMIF-Metabolic
  - EMIF-AD

- 56 partners from 14 European countries



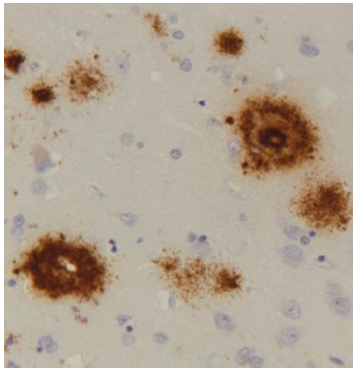
EMIF-AD

- Overall aim
  - Improve treatment opportunities for predementia AD by:
    - Discovery diagnostic and prognostic markers
    - Increased understanding AD pathophysiology

- Approach
  - Use existing data
    - Build infrastructure for data access and datasharing
  - Use extreme phenotypes as outcome
    - Amyloid positive vs amyloid negative

Alzheimer's disease

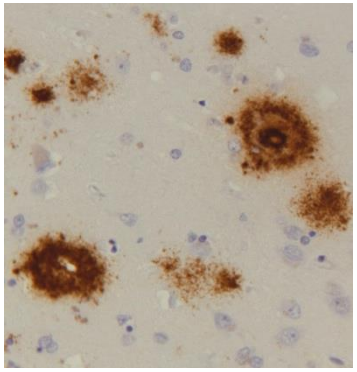
- Most common cause of dementia
- Starts with amyloid aggregation in the brain (plaques)



Plaques in brain

Alzheimer's disease

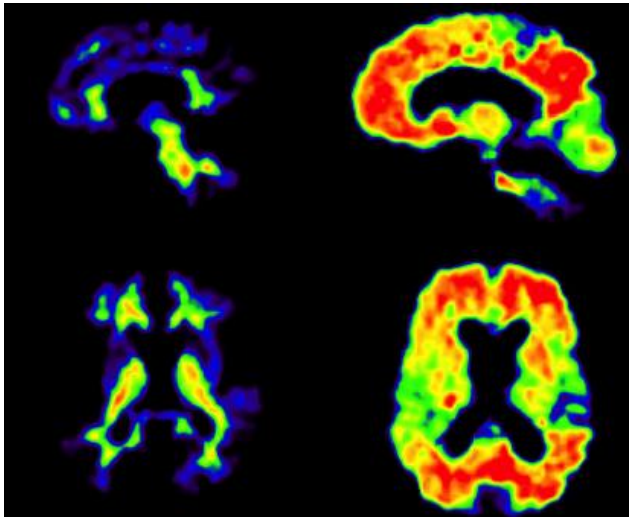
- Most common cause of dementia
- Starts with amyloid aggregation in the brain (plaques)



Plaques in brain

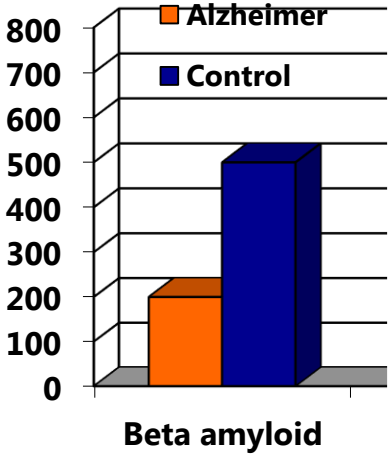
In-vivo amyloid measures

PET scan



Normal Alzheimer

Lumbar puncture



Amyloid decreased in CSF

Needs in Alzheimer's disease

- Large clinical datasets for:
  - Studies on etiology
  - Studies on prevalence and course disease
  - Selection of subjects for other studies
  - Monitoring treatment effects

- Type of data
  - EHR
  - Research cohorts
    - Clinical based
    - Population-based
    - Speciality groups

Researcher incentives for data sharing

- Valid research question
  - Can not be addressed by own data
- Acknowledgement in publication
- Nice to have
  - Funding
  - Access to pooled data for other analysis

Technical needs

- Find data:
  1. EMIF Catalogue
  2. EMIF-AD Participant selection tool (PST)

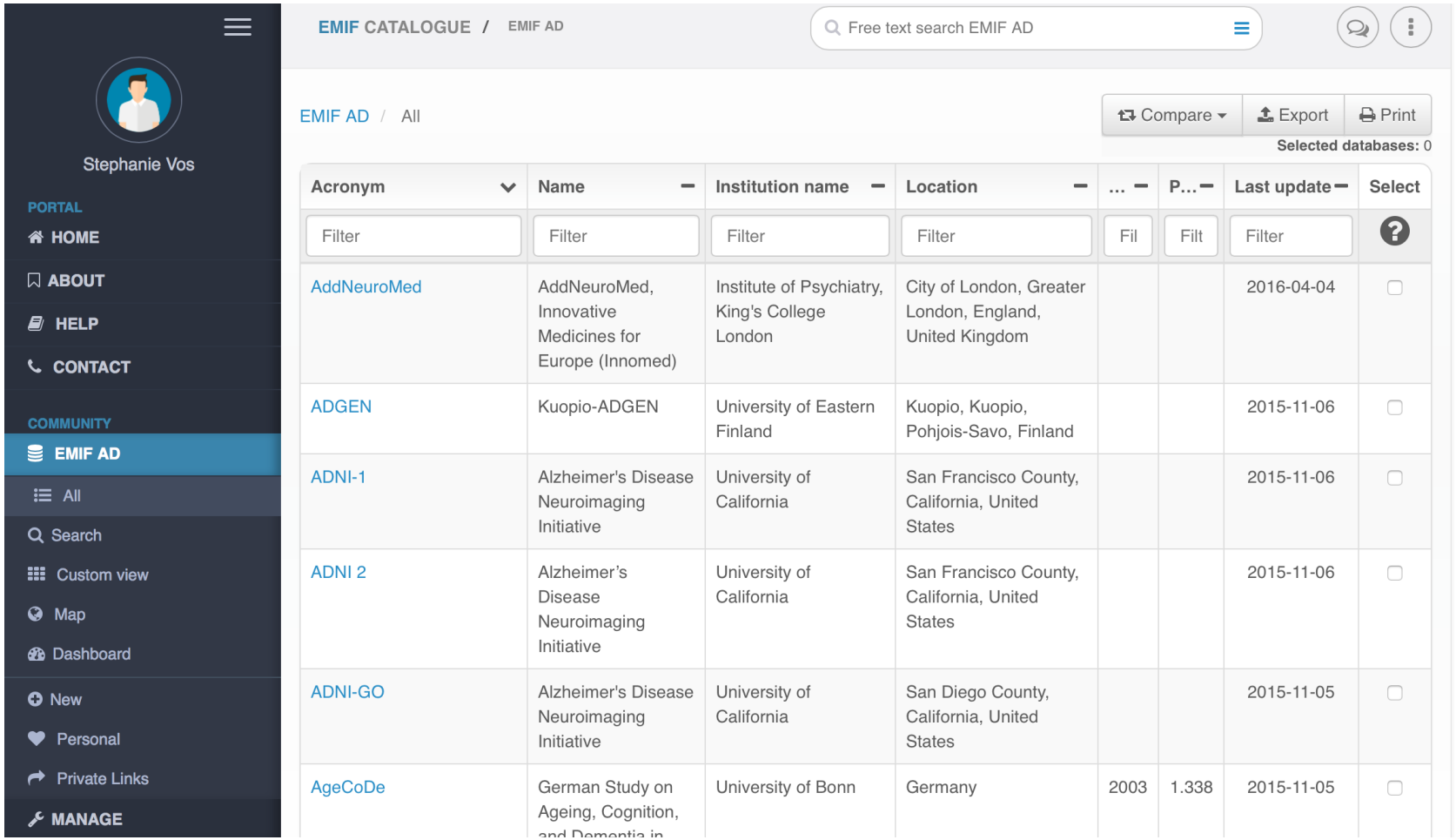
- Harmonise data:
  3. EMIF data model

- Access and analyse research data:
  4. TranSMART data platform

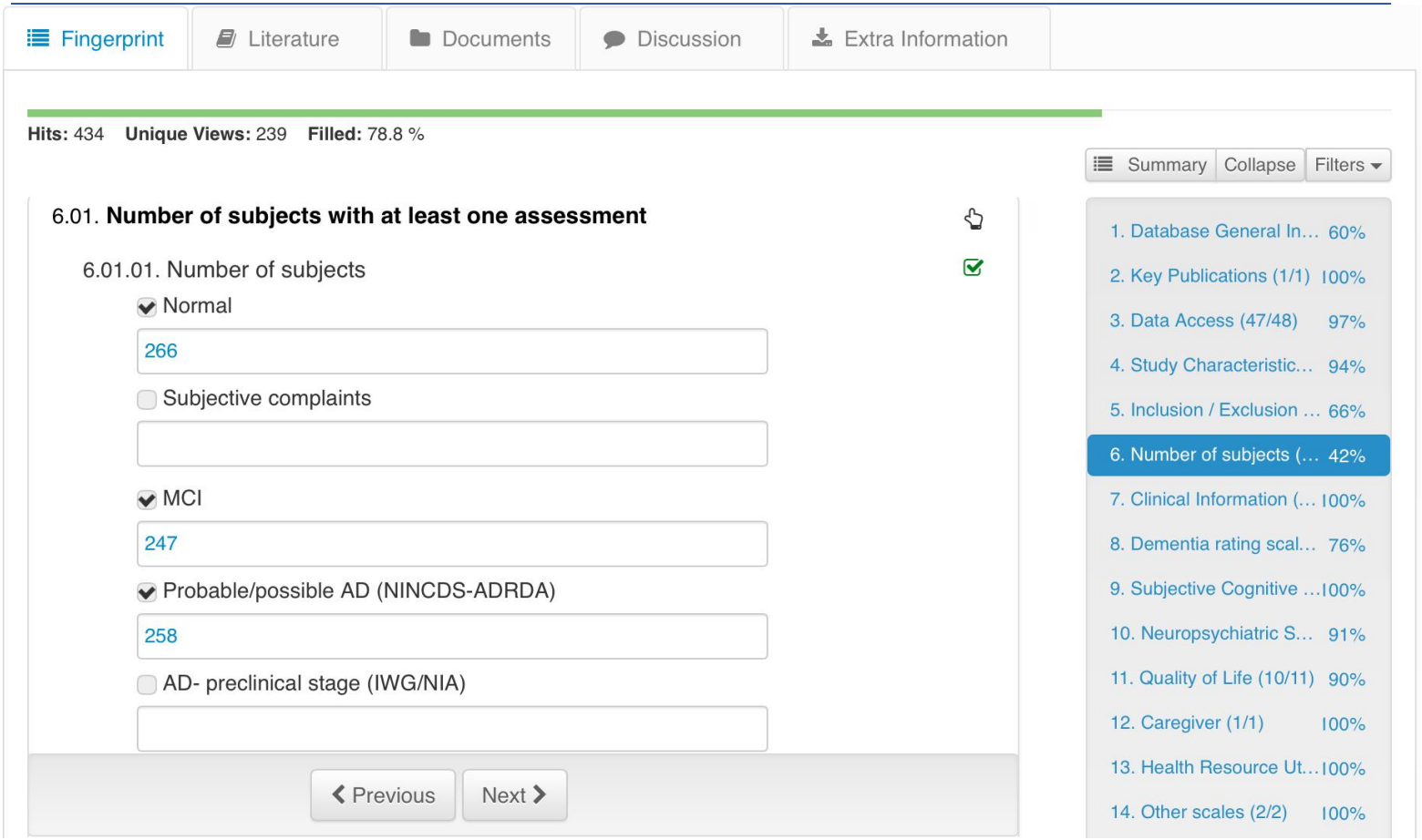
- Access and analyse EHR data:
  5. Jerboa and Octopus

1. EMIF catalogue (emif-catalogue.eu)

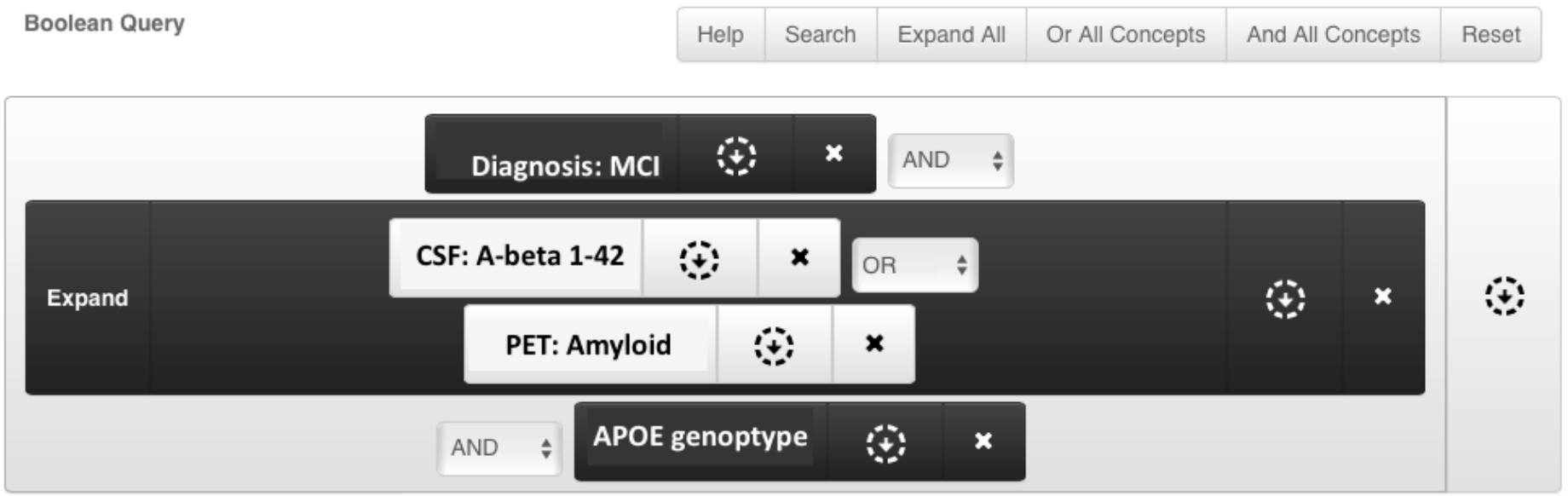
Meta-data of research cohorts and EHR datasets



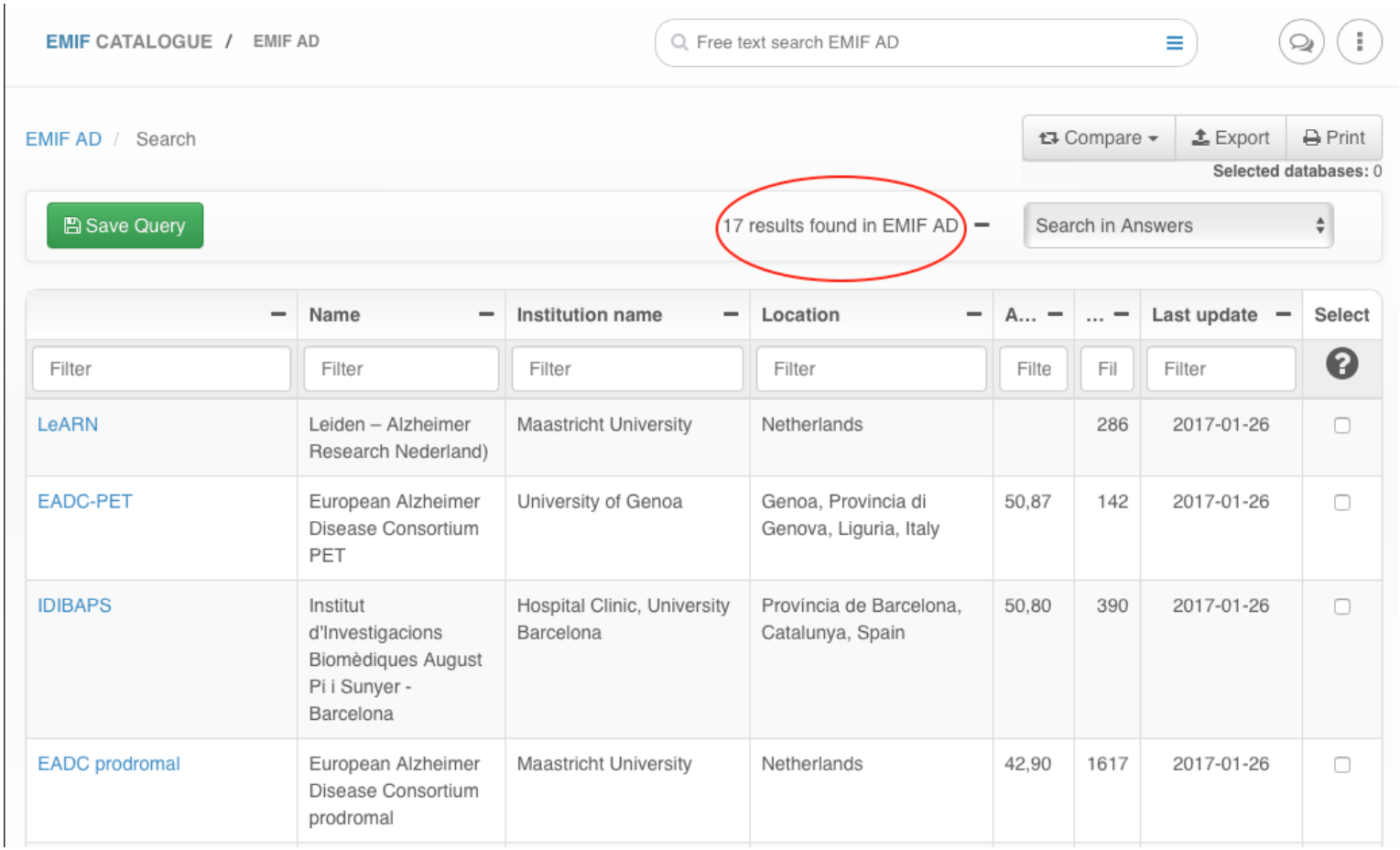
1. EMIF catalogue data entry



1. EMIF catalogue search



1. EMIF catalogue search



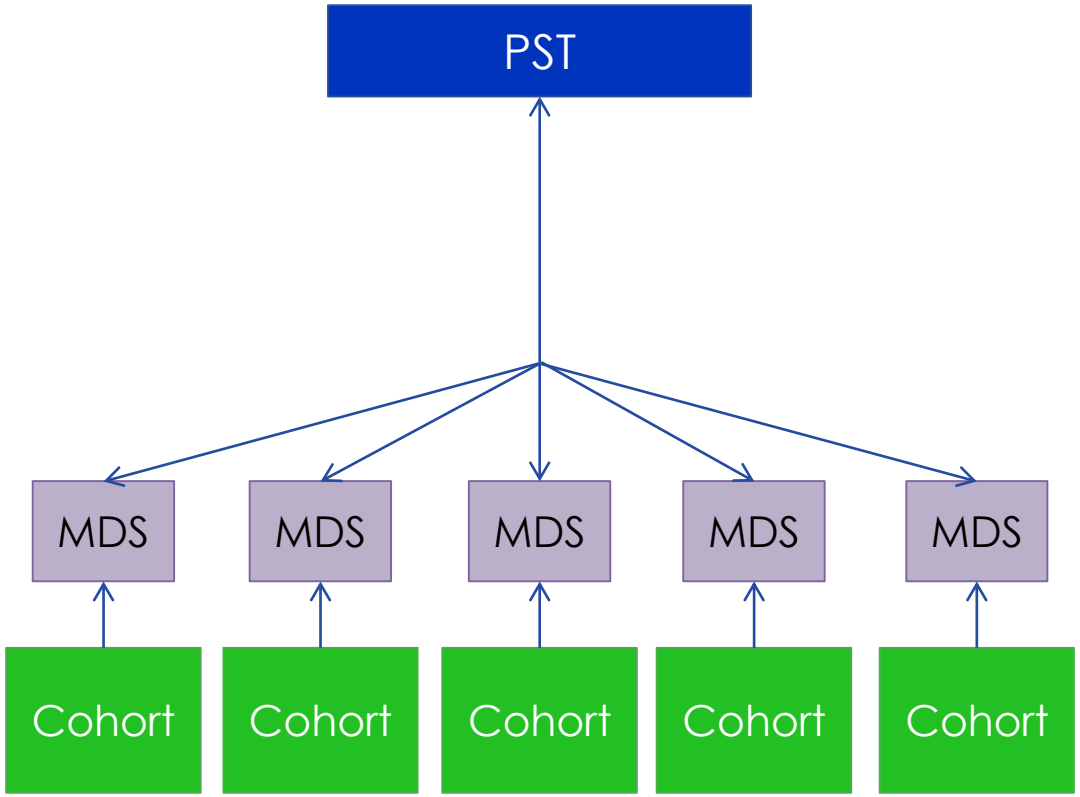
	Name	Institution name	Location	A...	...	Last update	Select
Filter	Filter	Filter	Filter	Filte	Fil	Filter	
LeARN	Leiden – Alzheimer Research Nederland)	Maastricht University	Netherlands		286	2017-01-26	
EADC-PET	European Alzheimer Disease Consortium PET	University of Genoa	Genoa, Provincia di Genova, Liguria, Italy	50,87	142	2017-01-26	
IDIBAPS	Institut d'Investigacions Biomèdiques August Pi i Sunyer - Barcelona	Hospital Clinic, University Barcelona	Província de Barcelona, Catalunya, Spain	50,80	390	2017-01-26	
EADC prodromal	European Alzheimer Disease Consortium prodromal	Maastricht University	Netherlands	42,90	1617	2017-01-26	

1. EMIF catalogue cohorts

	Number of cohorts
Clinical	31
Population	7
Trial	1
Other	5

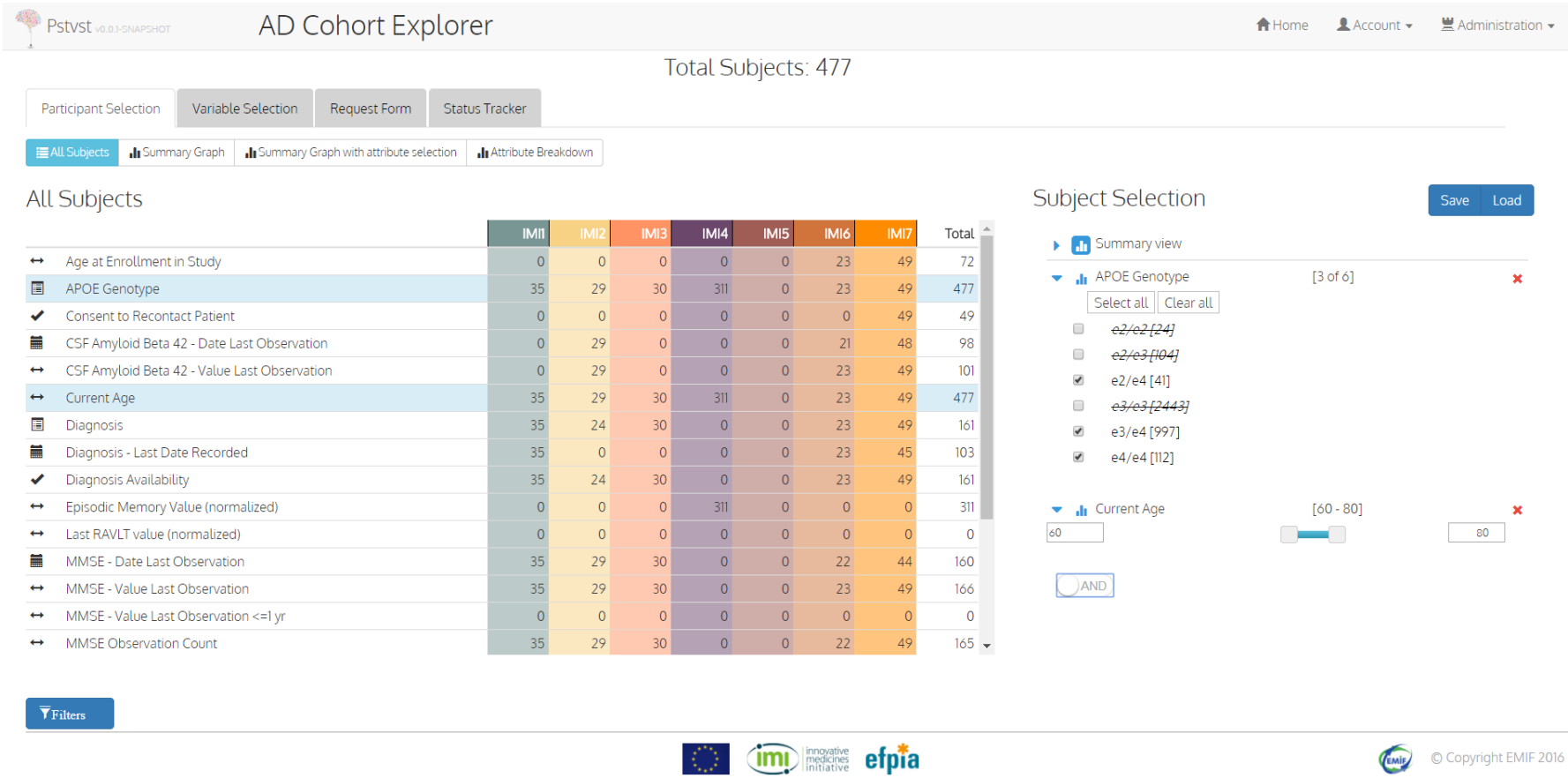
	Number of subjects
Normal cognition	49.972
Subjective cognitive complaints	4.416
Mild cognitive impairment	10.843
Alzheimer's disease dementia	9.949
Other dementia	2.453
	Total N=77.633

2. Participant selection tool

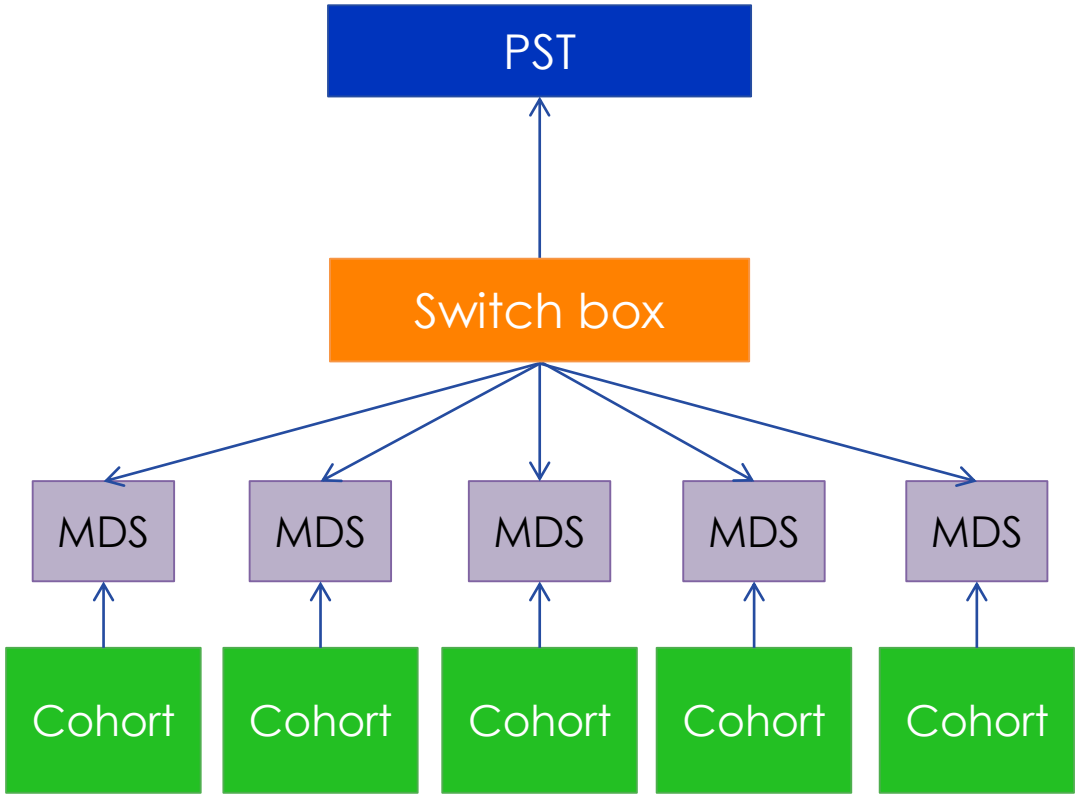


MDS=minimal dataset

2. Participant selection tool

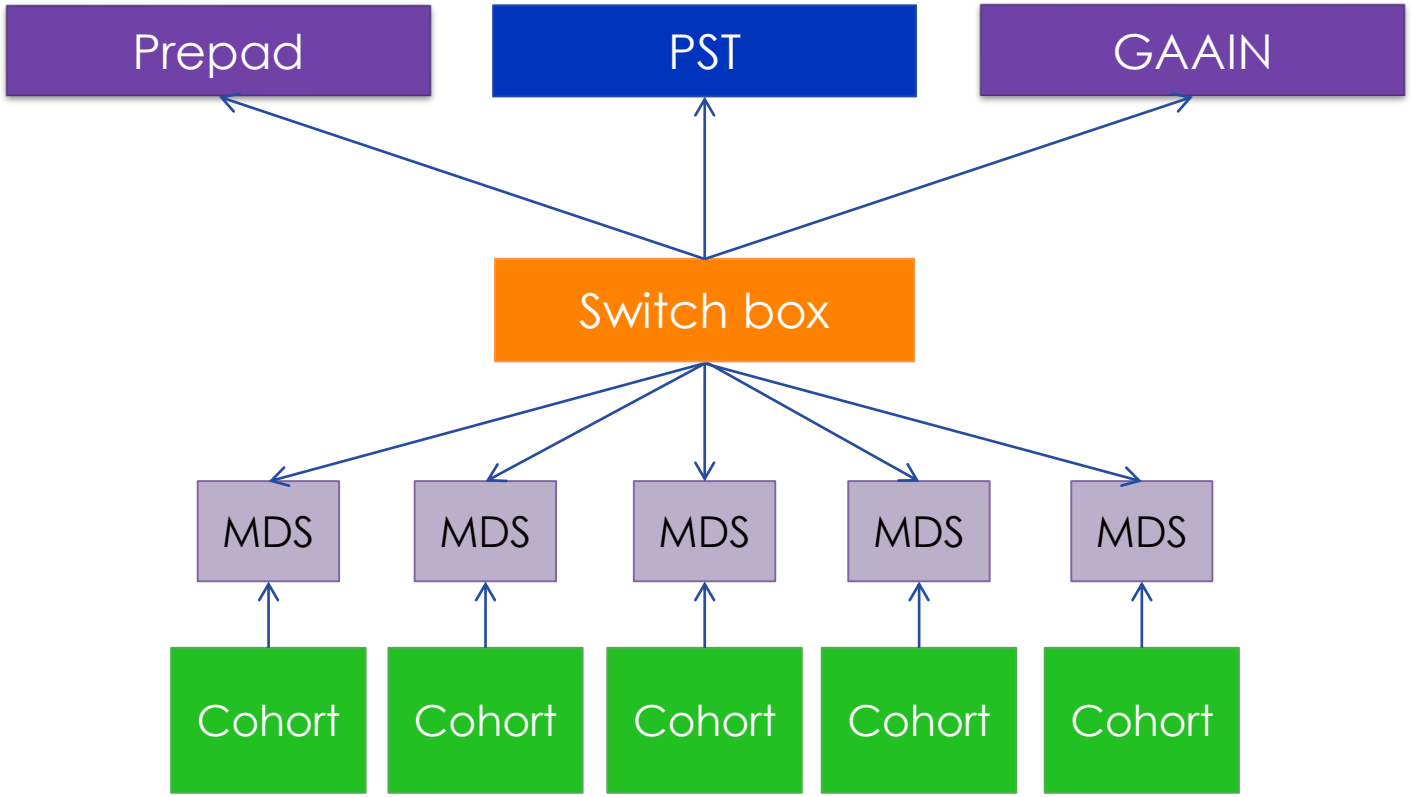


2. Participant selection tool



MDS=minimal dataset

2. Participant selection tool



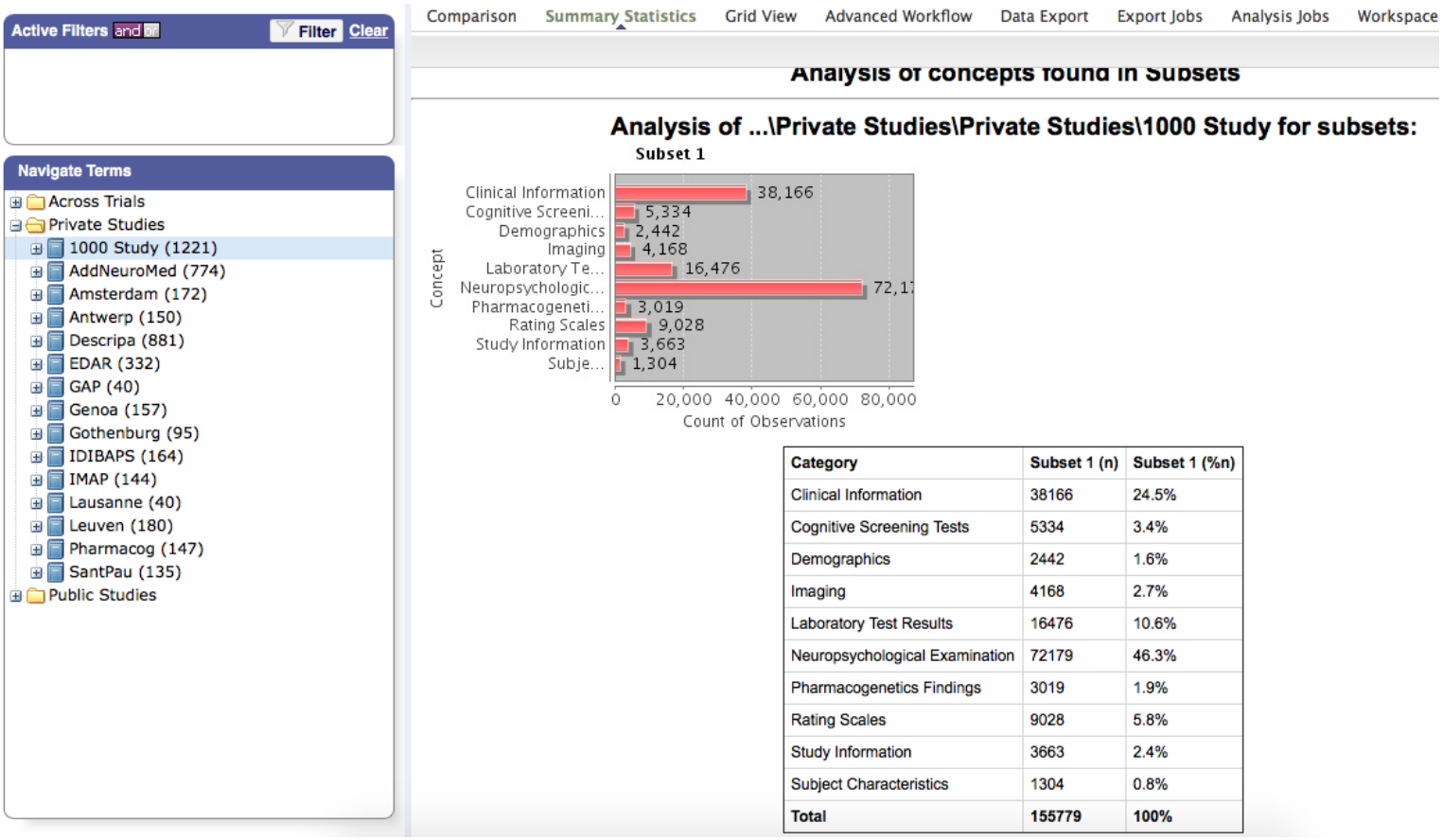
MDS=minimal dataset

3. Data harmonisation

- Research cohorts
  - EMIF-AD common data model
  - CDISC compliant
  - Minimal dataset of 50 variables

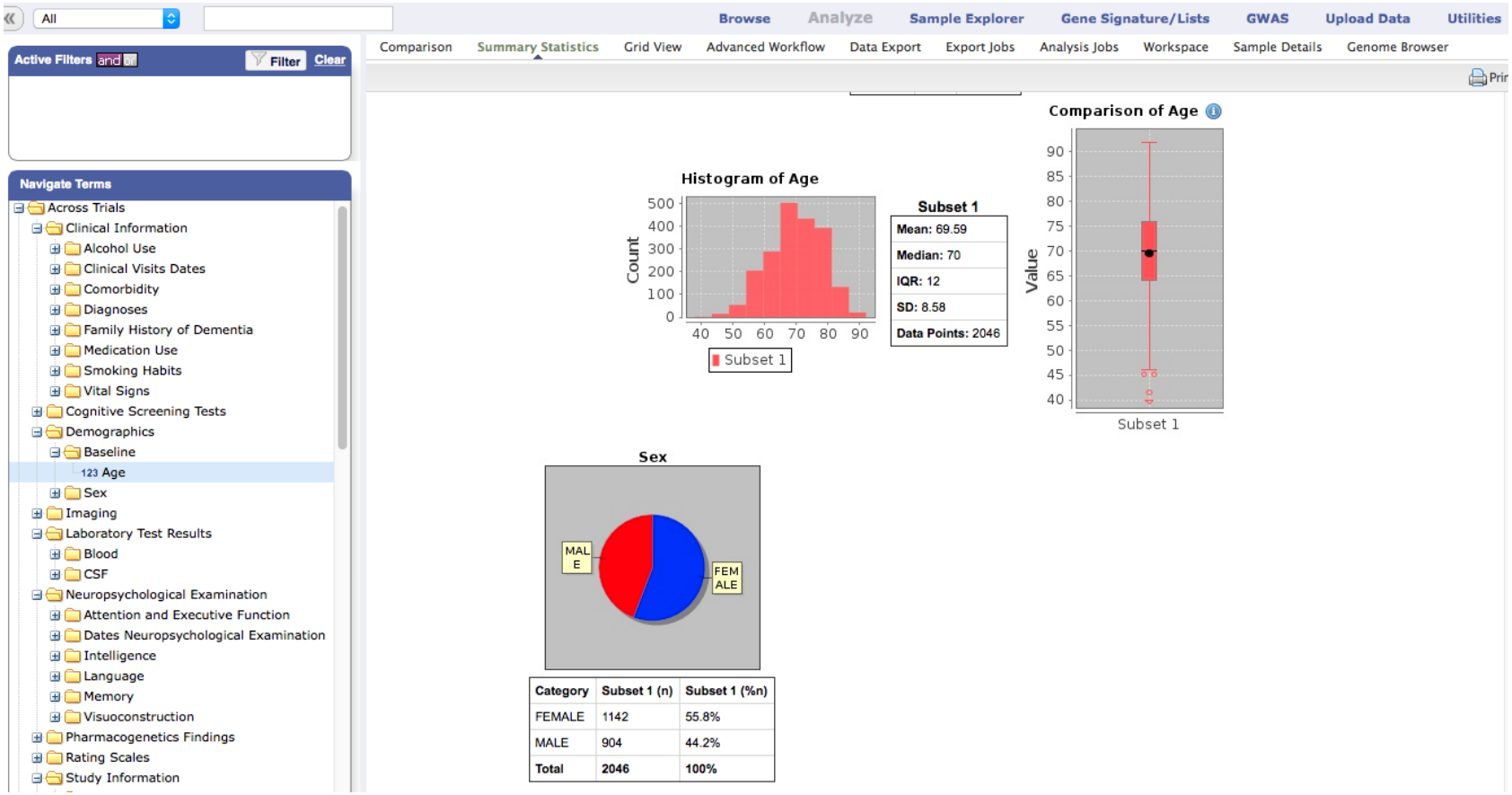
- EHR
  - OMOP common data model

4. EMIF-AD tranSMART dataplatform



Category	Subset 1 (n)	Subset 1 (%n)
Clinical Information	38166	24.5%
Cognitive Screening Tests	5334	3.4%
Demographics	2442	1.6%
Imaging	4168	2.7%
Laboratory Test Results	16476	10.6%
Neuropsychological Examination	72179	46.3%
Pharmacogenetics Findings	3019	1.9%
Rating Scales	9028	5.8%
Study Information	3663	2.4%
Subject Characteristics	1304	0.8%
Total	**155779**	**100%**

4. EMIF-AD tranSMART dataplatform



4. TranSMART cohorts

14 cohorts	Total N=3423
AddNeuroMed	786
Amsterdam	172
Antwerp	150
CITA	40
Descripa	881
EDAR	332
Genoa	157
Gothenburg	95
IDIBAPS	164
IMAP	144
Lausanne	40
Leuven	180
Pharmacog	147
Sant Pau	135

4. TranSMART minimal dataset

Demographics
- Age
- Gender
- Years of education

Clinical baseline information
- Diagnosis
- Functional impairment scale
- Depression scale
- Mini Mental State Examination
- Co-morbidities
- Medication use
- Date of baseline visit

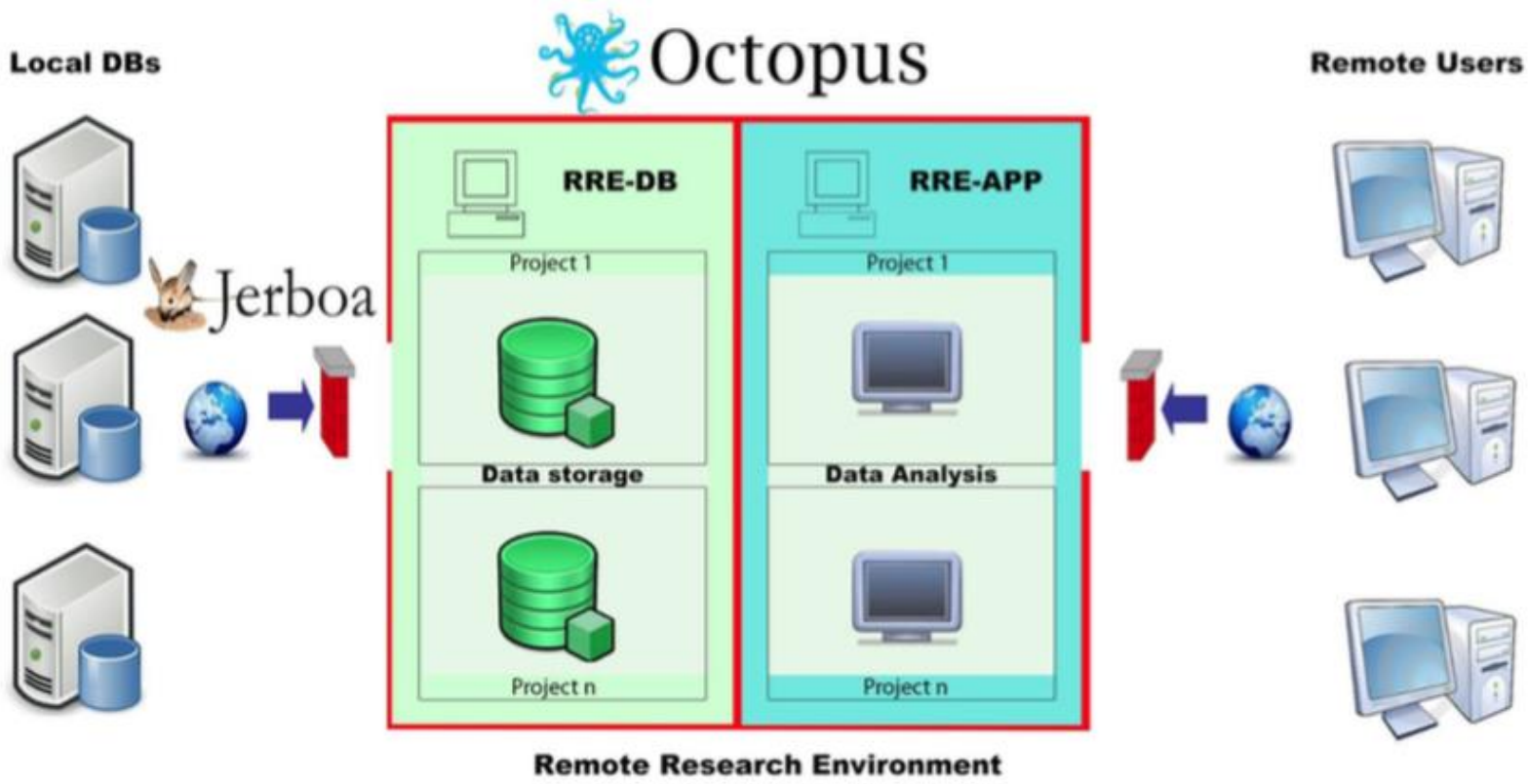
Baseline Neuropsychological raw scores and z-scores
- Memory test
- Language test
- Attention/Executive functioning test
- Visuoconstruction test
- Date of Neuropsychological examination

Amyloid measure
- Amyloid measure assessed by CSF or PET
- Date of amyloid assessment
- Cut-off used to define abnormality

MRI measure
- Measure of hippocampal volume or medial temporal atrophy
- Date of MRI assessment
- Cut-off used to define abnormality

Clinical follow-up data
- Last diagnosis
- Date of last clinical visit
- MMSE at each follow-up
- Date of MMSE at each follow-up
- Neuropsychological test scores at each follow-up
- Date of Neuropsychological test scores at each follow-up

5. EHR data access



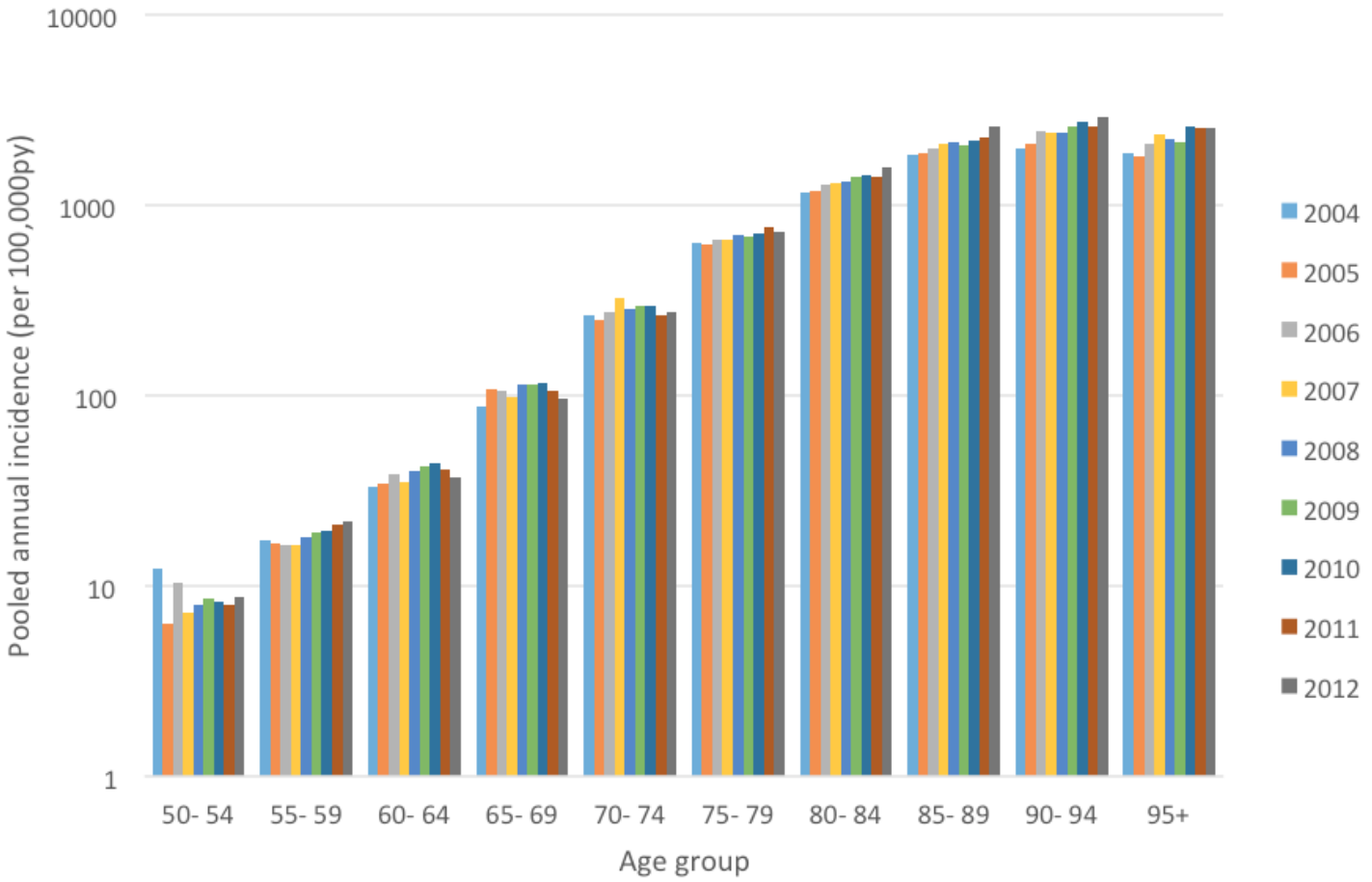
5. EMIF-associated EHR datasets

Database name	Setting	Total
THIN	General practitioner	12 million
IPCI	General practitioner	2.8 million
HSD	General practitioner	2.3 million
AUH	Hospital	2.3 million
IMASIS	Hospital	> 1.5 million
GePaRD	Health insurance data	17 million
ARS	Health insurance data	5 million
PHARMO	Drug prescriptions	10 million
EGCUT	Biobank	52,000
TOTAL		**52 million**

Examples of reuse data in AD

- Prevalence and incidence of dementia in EHR
- Prevalence of predementia AD in research cohorts
- Recruitment from existing cohorts
  - EMIF-AD biomarker discovery study
  - Preclin AD cohort

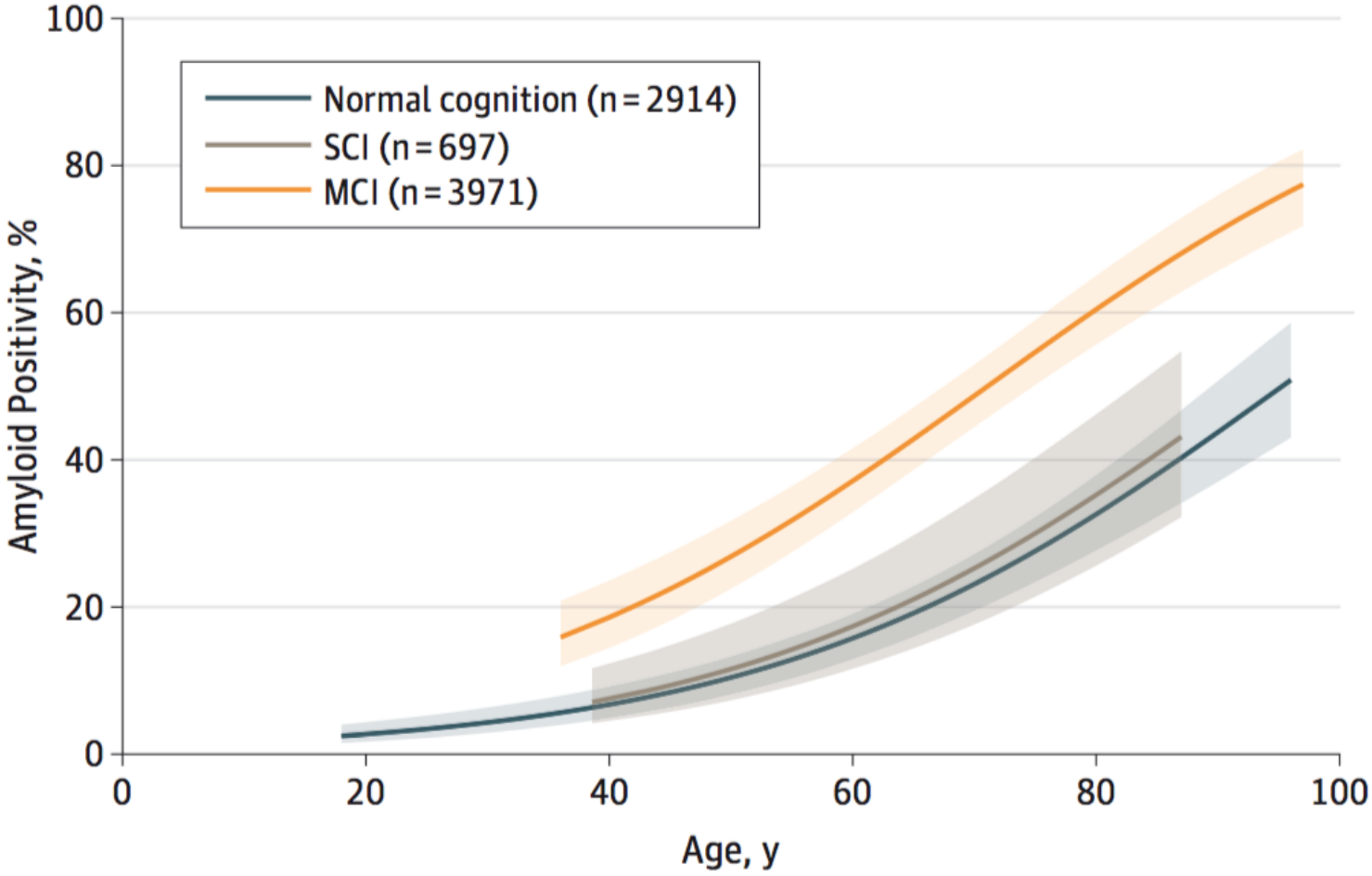
Incidence AD in EHR



Data from 6 European EHR datasets (n=25 million) with 138.000 dementia cases

Pereira et al Alz Dem 2017

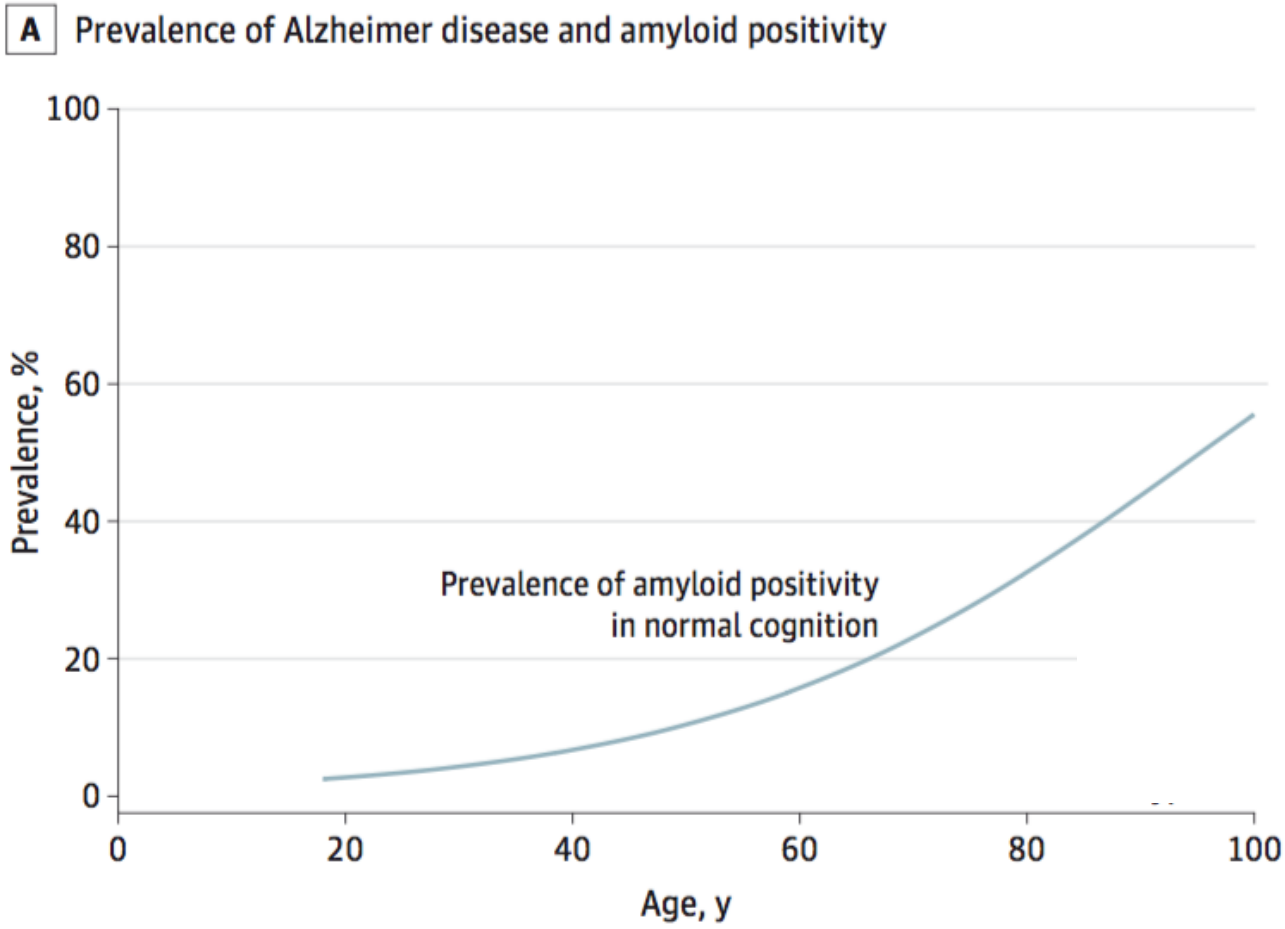
Prevalence predementia AD



Data from 51 research cohorts (n=8000)

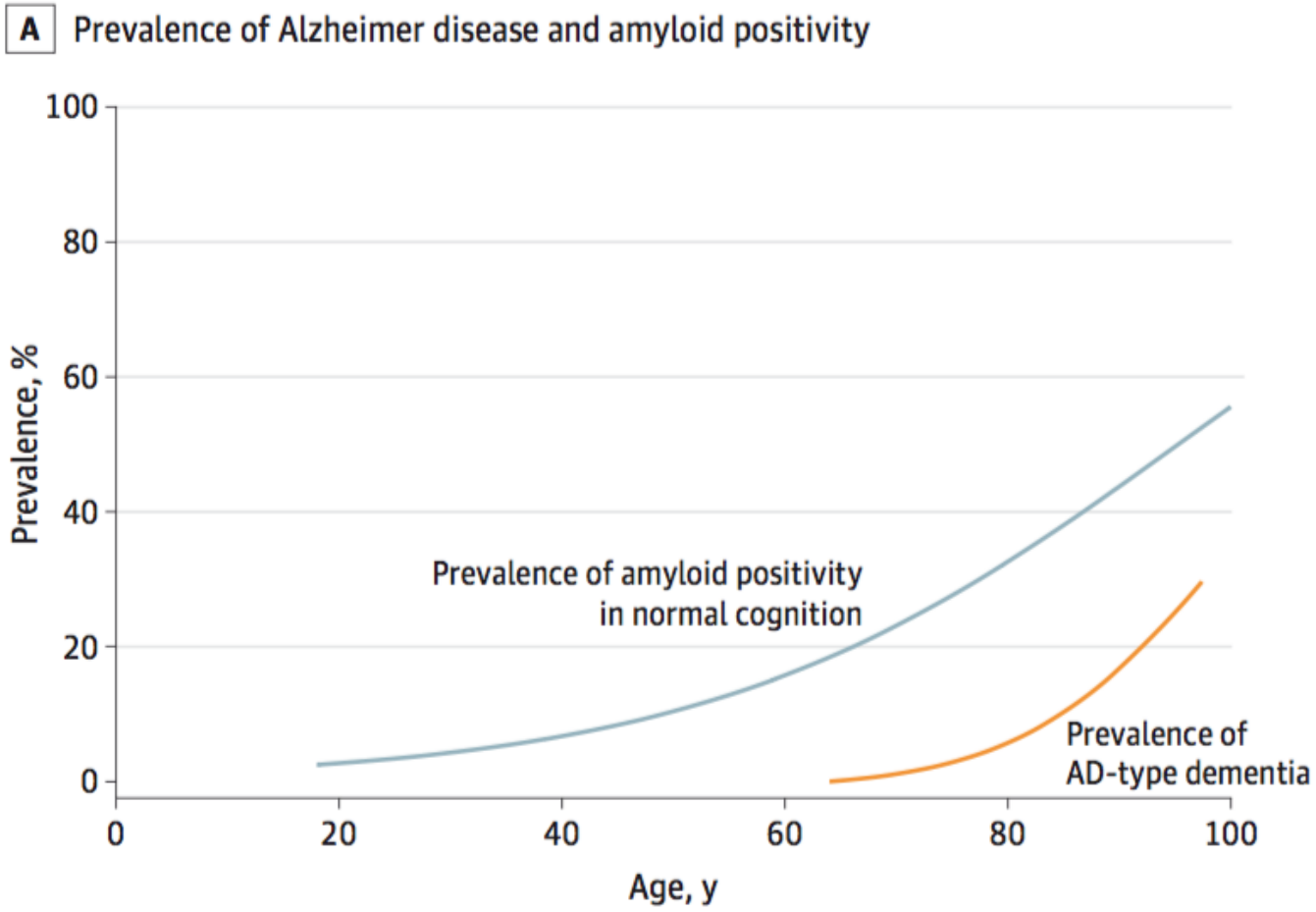
Jansen et al JAMA 2015

Example: Prevalence predementia AD



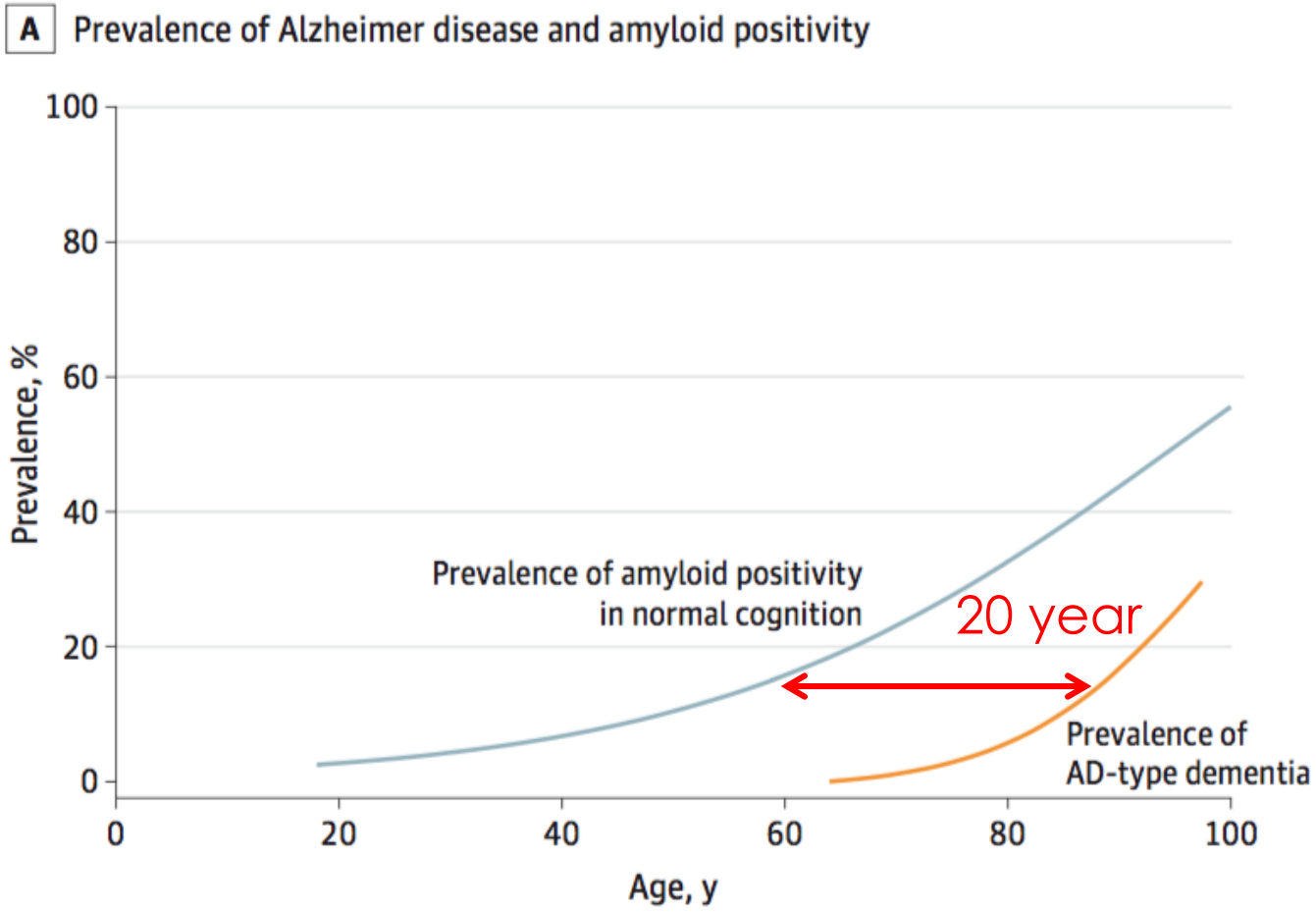
Jansen et al JAMA 2015

Example: Prevalence predementia AD



Jansen et al JAMA 2015

Example: Prevalence predementia AD



Jansen et al JAMA 2015

Recruitment existing cohorts:
Biomarker discovery study

- Aim: find novel diagnostic and prognostic markers for predementia AD using existing data and samples

- Steps
  - Identification of cohorts through EMIF catalogue
  - Set-up contracts
  - Data pooling in tranSMART, central sample storage

- Status
  - 1200 subjects enrolled
  - Analysis ongoing

Recruitment existing cohorts:
PreclinAD and 90+ studies

PreclinAD study (n=280)

90+ study (n=120)

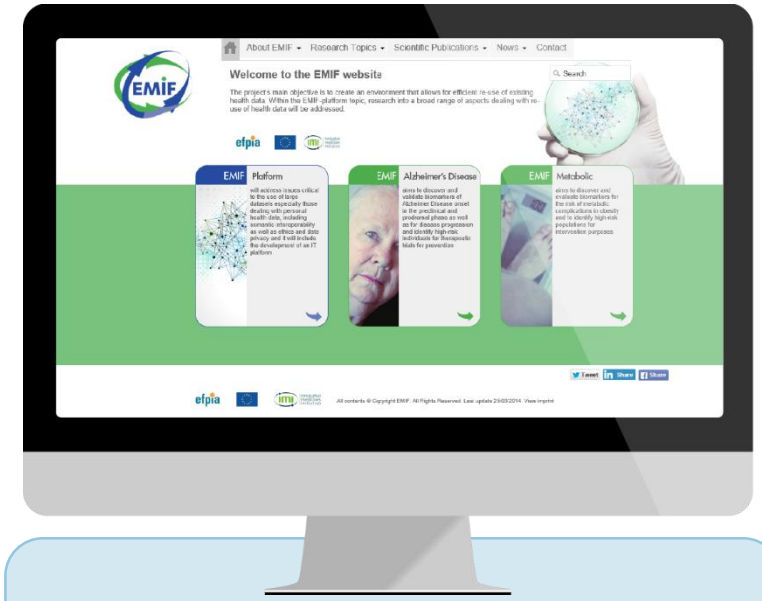
- Recruited from
  - Netherlands Twin Registry, Manchester and Newcastle Aging study, hospital settings

Acknowledgements

- Co-PI's: Simon Lovestone, Johannes Streffer
- WP1: Stephanie Vos, Isabelle Bos, Karl Herholz, Stephan Carter, Rainer Hinz, Elles Konijnenberg, Anouk Den Braber, Jori Tomassen
- WP2: Gerald Novak, Mark Gordon, Sebastiaan Engelborghs, Jana Podhorna, Gayan Perera, Rob Stewart, Elizabeth Baker, Alejo Nevado-Holgado, Wenbo Tang, Preciosa Coloma, Lisa Ford, Nienke Legdeur, Maryam Badissi, Christian Schultheis, Rosemary Abbot, Hilkka Liedes
- WP3: Alison Baird, Cristina Legido-Quigley, Richard Dobson, Stephen Newhouse, Sarah Westwood, Anna Myers, Stuart Snowden, Malcom Ward, Lars Bertram, Kristel Sleegers, Katie Lunnon, Henrik Zetterberg, Frederik Barkhof, Mara Ten Kate, Giovanni Frisoni, Alberto Redolfi,
- WP4: Piotr Lewczuk, Hilkka Soininen, Henrik Zetterberg, Kina Höglund, Maria Pikkarainen
- PI's data-cohorts: Philip Scheltens, Sebastiaan Engelborghs, Giovanni Frisoni, Rik Vandenberghe, José Luis Molinuevo, Anders Wallin, Alberto Lléo, Julius Popp, Pablo Martinez-Lage

More Information

- **EMIF general**
  - Bart Vannieuwenhuyse (bvannieu@its.jnj.com)
  - Simon Lovestone (simon.lovestone@psych.ox.ac.uk)
  - Johan van der Lei (j.vanderlei@erasmusmc.nl)
- **EMIF-Platform**
  - Johan van der Lei (j.vanderlei@erasmusmc.nl)
  - Nigel Hughes (nhughes@its.jnj.com)
- **EMIF-Metabolic**
  - Ulf Smith (ulf.smith@medic.gu.se)
  - Dawn Waterworth (dawn.m.waterworth@gsk.com)
- **EMIF-AD**
  - Pieter Jelle Visser (pj.visser@maastrichtuniversity.nl)
  - Johannes Streffer (jstreffe@its.jnj.com)



www.emif.eu
Register for our newsletter!

EMIF is operating under IMI Grant Agreement n°115372


EMIF


efpia
imi
innovative
medicines
initiative

EMIF Cross Topic

Peter Egger

RWE & Epidemiology, GSK

Research Use Cases – what have we learned?
The EMIF EHR-Platform Perspective

Joint i~HD/EMIF Meeting 21st-22nd September 2017- Madrid

Peter Egger, Glen James, Myriam Alexander
Epidemiology, GlaxoSmithKline

efpia

imi innovative medicines initiative

Overview

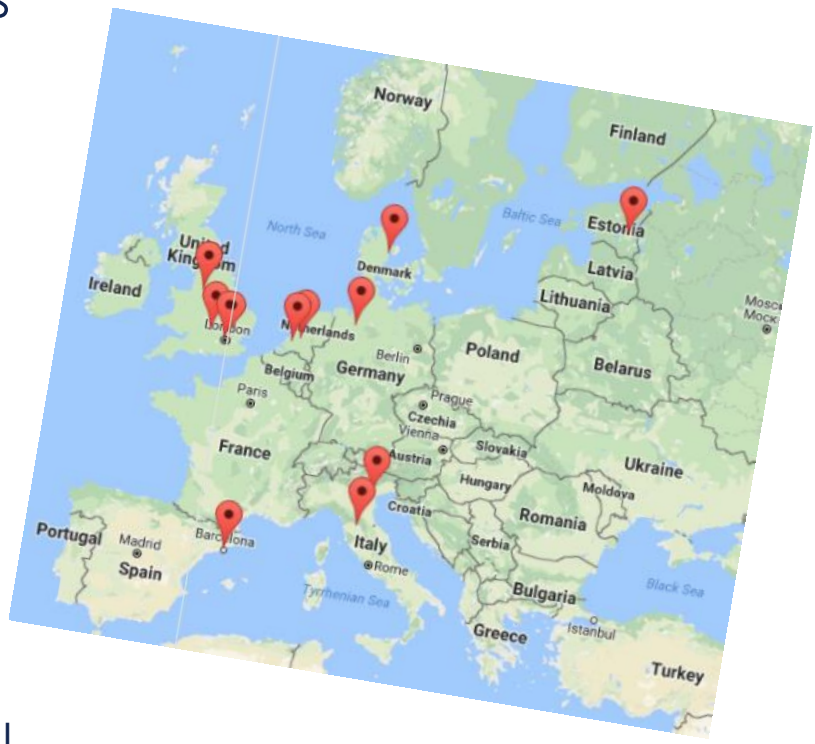
- Real world evidence to support healthcare
- The need for real world evidence from Europe
- What EMIF can offer
- Examples of studies conducted so far
- Summary and conclusion

Real world evidence to support healthcare

1. Determine unmet need and the value of intervention
2. Assess impact of health policy and resource allocation
3. Guide clinical development of new molecules
4. Evaluate the real world effects of medications

Real world evidence from across Europe

- Choice of different data sources
- Diversity
  - Geography
  - Healthcare systems and disease management
  - Type of healthcare data
- Large numbers
  - To evaluate rare occurrences
- Need for integrated data
  - Comprehensive patient medical records



What EMIF can offer

Research collaborations based on a wide network of data sources within a **Common Environment**

Standard formats and tools and consistent ways of working across the different data sources

Consistent quality of research
More efficient study execution
Greater familiarity with study results format
More reliable comparisons

A Common Environment for the federated data network

Catalogue

Data source Characteristics

- Size
- Information content

Key dashboards

- Patient demographics
- Key clinical data

Open to all and free

Data Query

- Simple – numbers of patients only
- Fast & low cost
- Pre-approved

Full Study

Study execution – common processes

- contracting
- protocol, rev & approve
- semantic harmonisation
- data extraction
- analysis environment

Standard Modules

- Incid/prev
- Patient profile
- Treatment patterns
- Resource utilisation

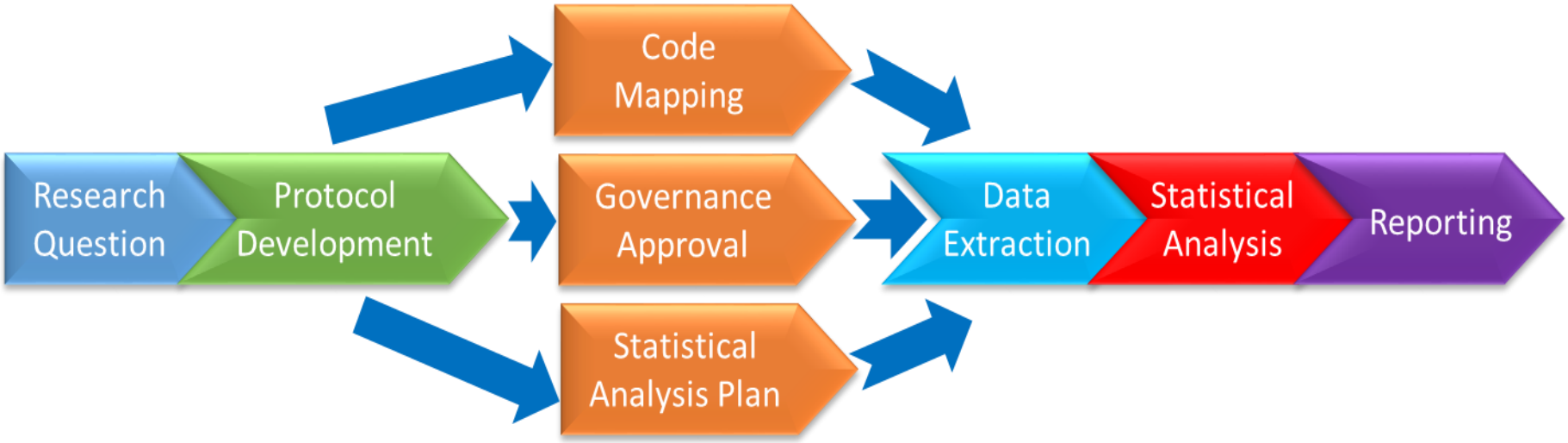
Bespoke Studies

- Disease natural history
- Drug effectiveness
- Drug safety

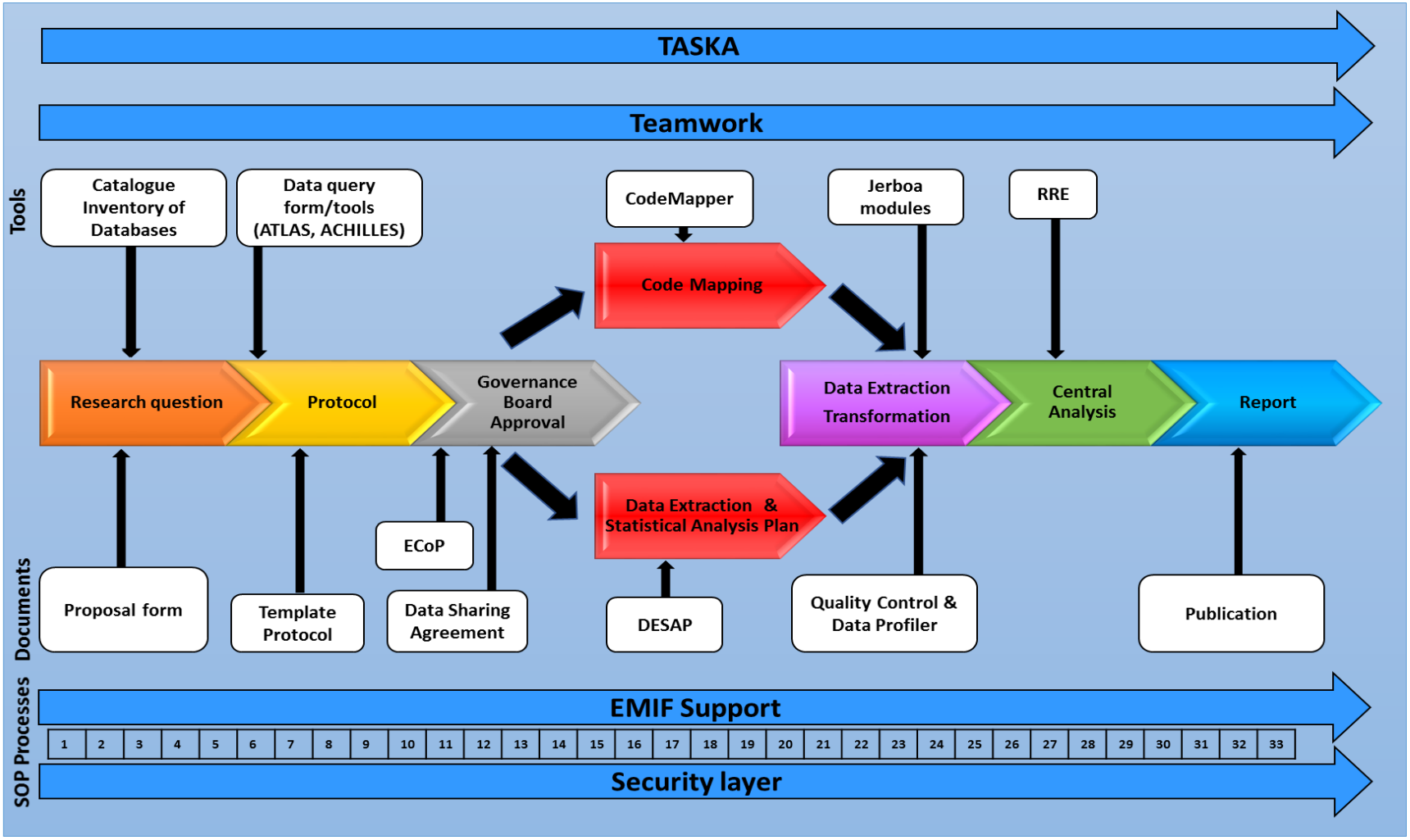
efpia

imi innovative medicines initiative

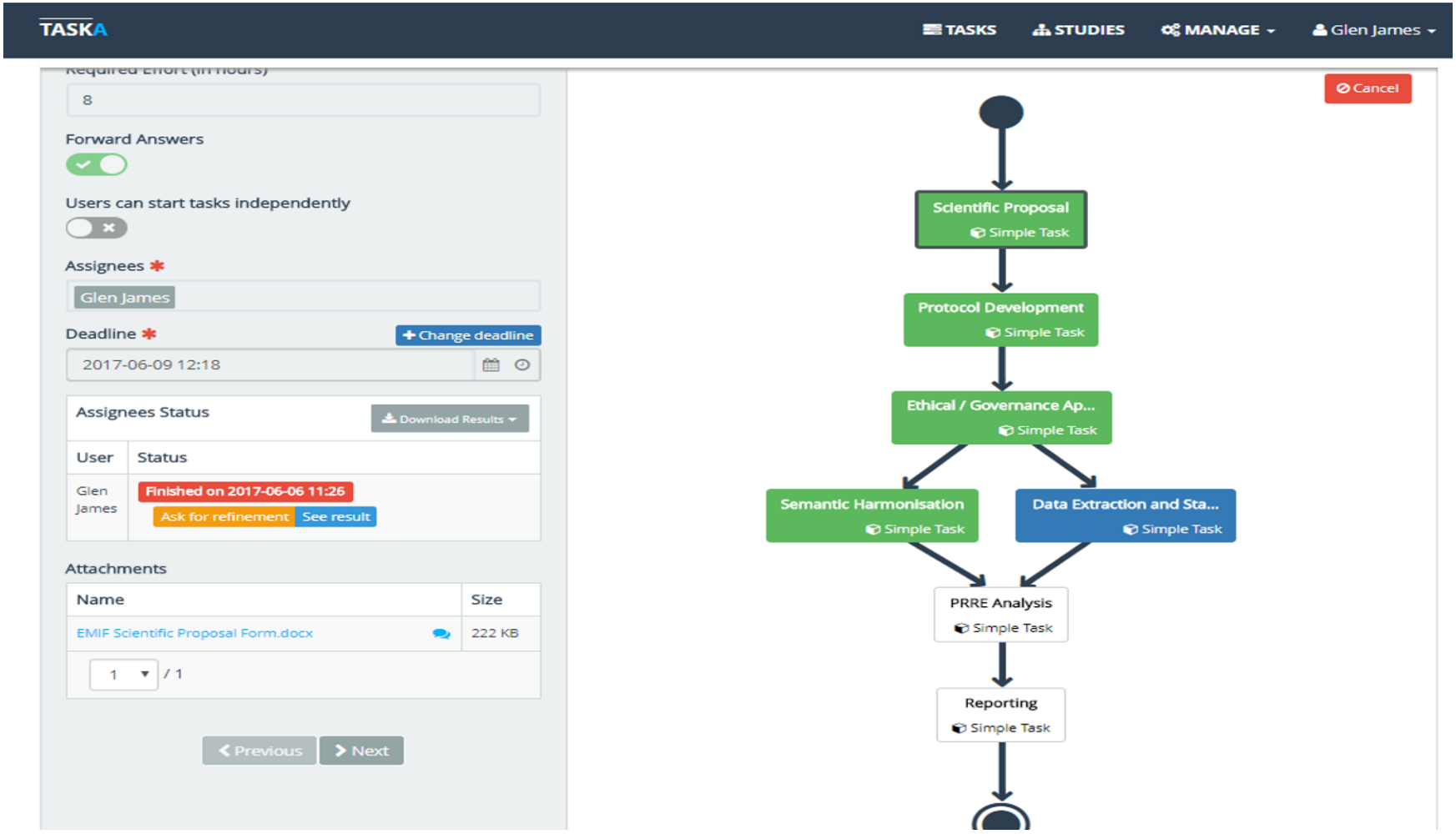
Roadmap for study execution



Detailed EHR Platform Roadmap



EMIF Templates & TASKA



CodeMapper

Heart failure

CODE MAPPER

ADVANCE

Case definition | *Mapping* | History

114 concepts*

Filter

Modify 0 selected concepts*

Delete* | Broader* | Narrower* | Suggest* | Tags* | Codes*

Search and add concept*

Query | Search

Operate on mapping

Coding systems* | Save* | Download* | Discard*

Concept	ICD10	ICD10CM	ICD9CM	ICPC2EENG	ICPC2P	MTHICD9	RCD2
Heart failure	Heart failure **I50** Heart failure, unspecified **I50.9**	Heart failure **I50** Heart failure, unspecified **I50.9**	Heart failure **428** Heart failure, unspecified **428.9**	Heart failure **K77**	Weakness;heart **K29019** Failure;cardiac	Cardiac failure NOS **428.9**	Heart failure **G58..** Heart failure NOS **G58z.**
Cor pulmonale	Pulmonary heart disease, unspecified **I27.9**	Cor pulmonale NOS **I27.81** Pulmonary heart		Pulmonary heart disease **K82**	Disease;heart;pulmonary **K82002** Cor pulmonale **K82003**	Cor pulmonale NOS **416.9**	
Congestive heart failure	Congestive heart failure **I50.0**	Congestive heart disease **I50.9**	Congestive heart failure, unspecified **428.0**		Failure;congestive cardiac **K77002**	Congestive heart disease **428.0**	Congestive heart failure **G580.**
Left-Sided Heart Failure	Left ventricular failure **I50.1**	Left ventricular failure **I50.1**	Left heart failure **428.1**		Failure;ventricular;left **K77008**	Left ventricular failure **428.1**	Left ventricular failure **G581.**
Hypertensive heart and renal disease with renal failure	Hypertensive heart and renal disease with renal failure **I13.1**		Hypertensive heart and chronic kidney disease, unspecified, without heart failure			Hypertensive heart and renal disease, unspecified, with renal failure **404.92**	Hypertn hrt&ren dis+renal fail **G233.**
Hypertensive heart and renal disease with both	Hypertensive heart and renal disease		Hypertensive heart and chronic kidney			Hypertensive heart and renal disease	Hyp ht&ren d+both(con)h&r fail **G234.**

efpia | innovative medicines initiative

Examples of pilot research projects (UseCases)

Selected pilot projects

Use Case	Title	Progress
6	**Dementia prevalence and incidence in a federation of European Electronic Health Record (EHR) databases.**	**Complete -** https://www.ncbi.nlm.nih.gov/pubmed/28734783
9	BMI and the risk of cardiovascular disease and all-cause mortality in European electronic medical records databases.	**Analysis ongoing**
10	Association of non-alcoholic fatty liver disease with cardiovascular and liver morbidity in electronic health record databases.	**Publication in draft**
11	**Dementia: vascular and metabolic risk factors**	**Publication in draft**
13	Treatment pathway analysis: An evaluation of treatment patterns and drug utilisation amongst cases with incident dementia in Electronic Health Records databases available in the European Medical Information Framework	**Analysis ongoing**
14	A nested case-control study of prior history of non-alcoholic fatty liver disease in demented and cognitively impaired individuals matched to healthy controls in European health records data.	**Governance Approval**
15	Utilisation of healthcare data to identify sub-types of heart failure patients based on clinical and/or molecular phenotypes	**Data Extraction**
16	An exploratory phenome wide association study linking asthma & liver disease single nucleotide polymorphisms and electronic health records from the Estonian Genome Centre at the University of Tartu Database	**Governance Approval**

Key data sources

Database name	Total number of subjects
AUH - Denmark, hospital (Aarhus) & prescriptions	2.3 million
THIN - UK, primary care	12 million
IPCI - Netherlands, primary care	2.8 million
HSD - Italy, primary care	2.3 million
IMASIS - Spain, Barcelona, hospital	> 1.5 million
PEDIANET - Italy, pediatrics	0.4 million
PHARMO - Netherlands, linked databases	10 million
SIDIAP - Spain, Catalonia, primary care	6 million
ARS - Italy, Tuscany, hospital & prescriptions	5 million
EGCUT - Estonia, total healthcare & biobank	52,000

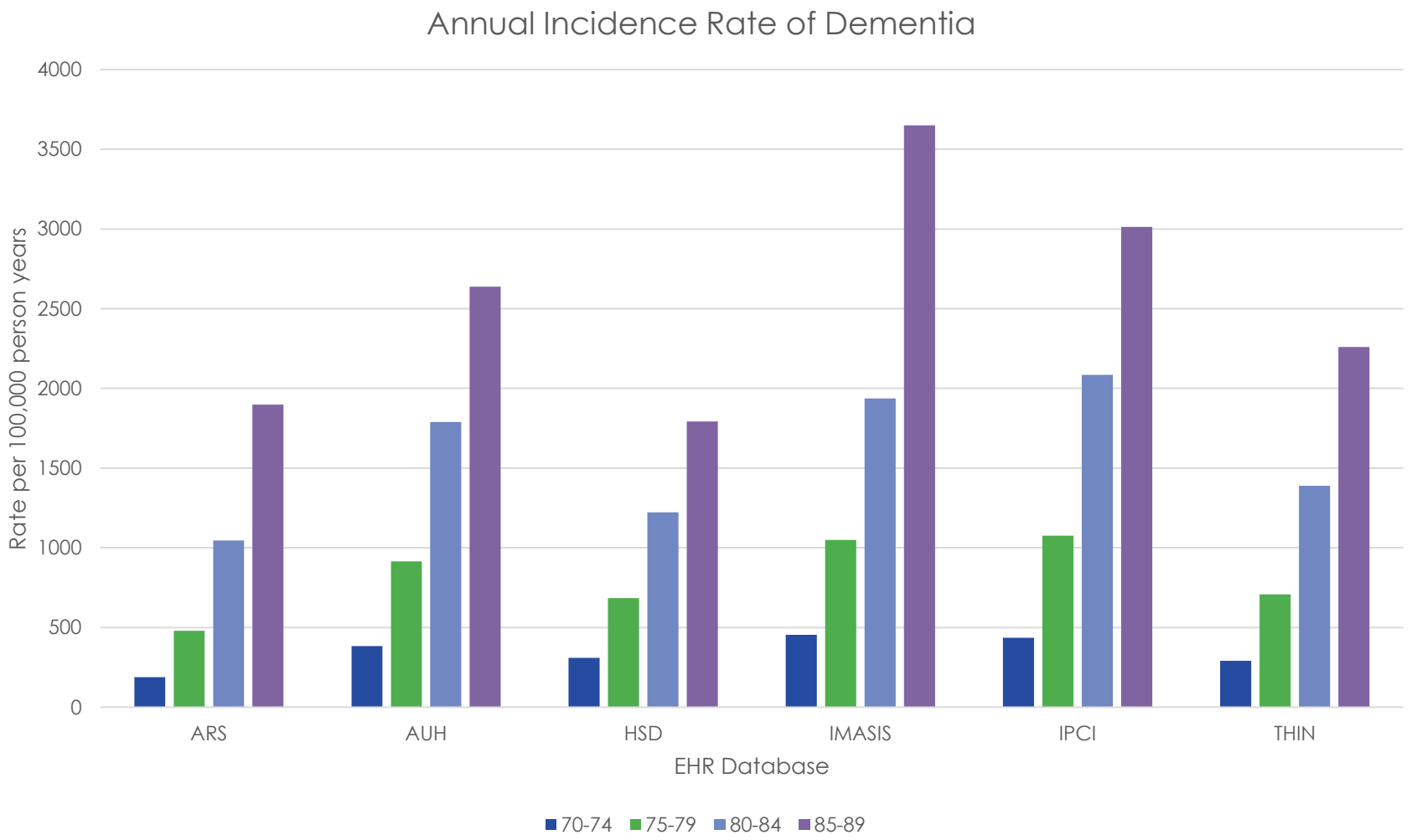
UC6: Dementia prevalence & incidence in a federation of European EHR databases: The EMIF resource.

Perera G, Pedersen L, Ansel D, Alexander M, Arrighi HM, Avillach P, Foskett N, Gini R, Gordon MF, Gungabissoon U, Mayer MA, Novak G, Rijnbeek P, Trifirò G, van der Lei J, Visser PJ, Stewart R.

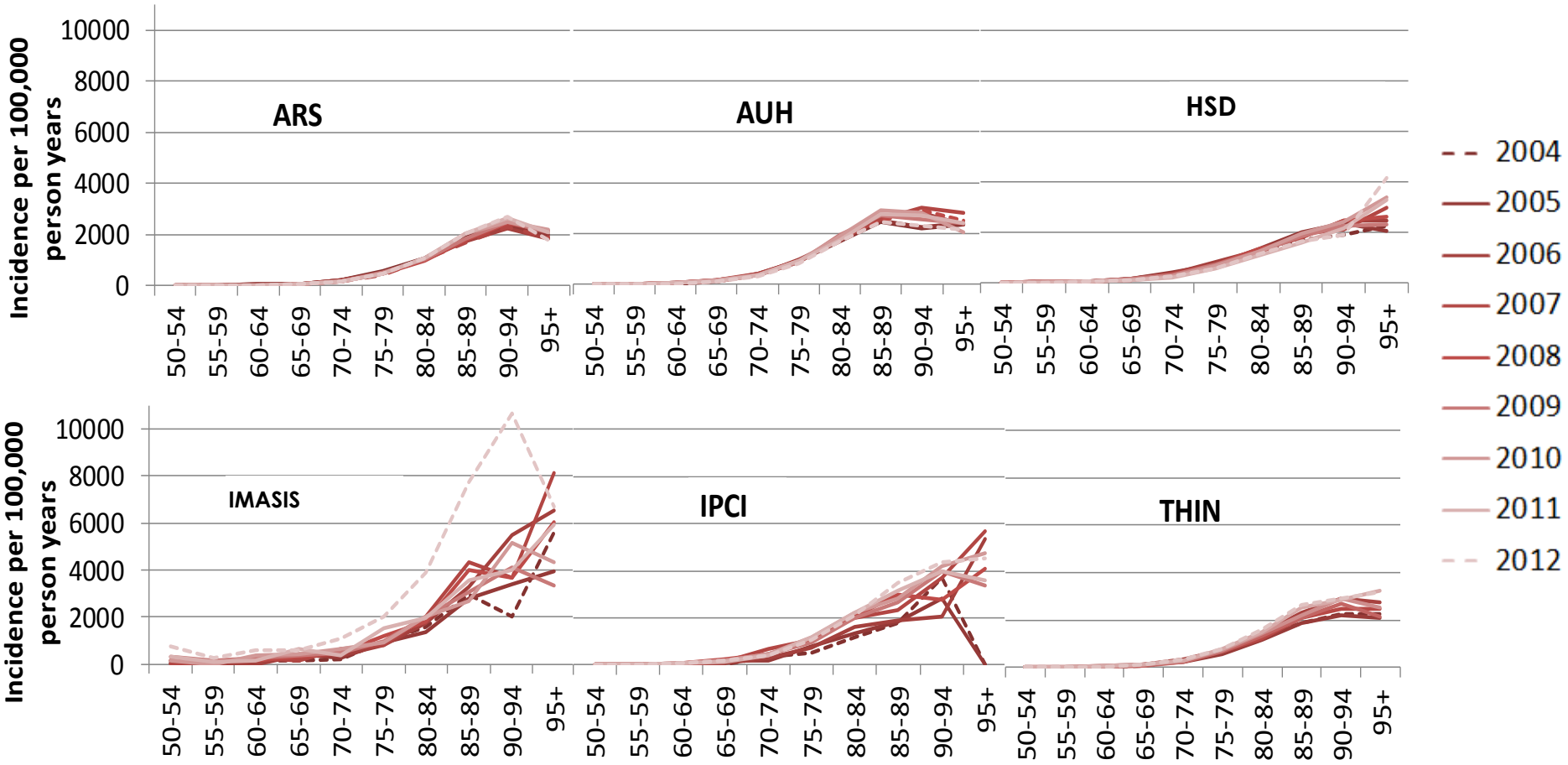
Alzheimer's and Dementia (2017), 1-10

- 6 EHR databases analysed (ARS, AUH, IPCI, HSD, IMASIS, THIN)
- Identified 139,000 dementia cases from an overall total of 25 million persons from 2004 to 2012
- Results lower than in the published literature but similar secular trends and patterns over age

Incidence of Dementia

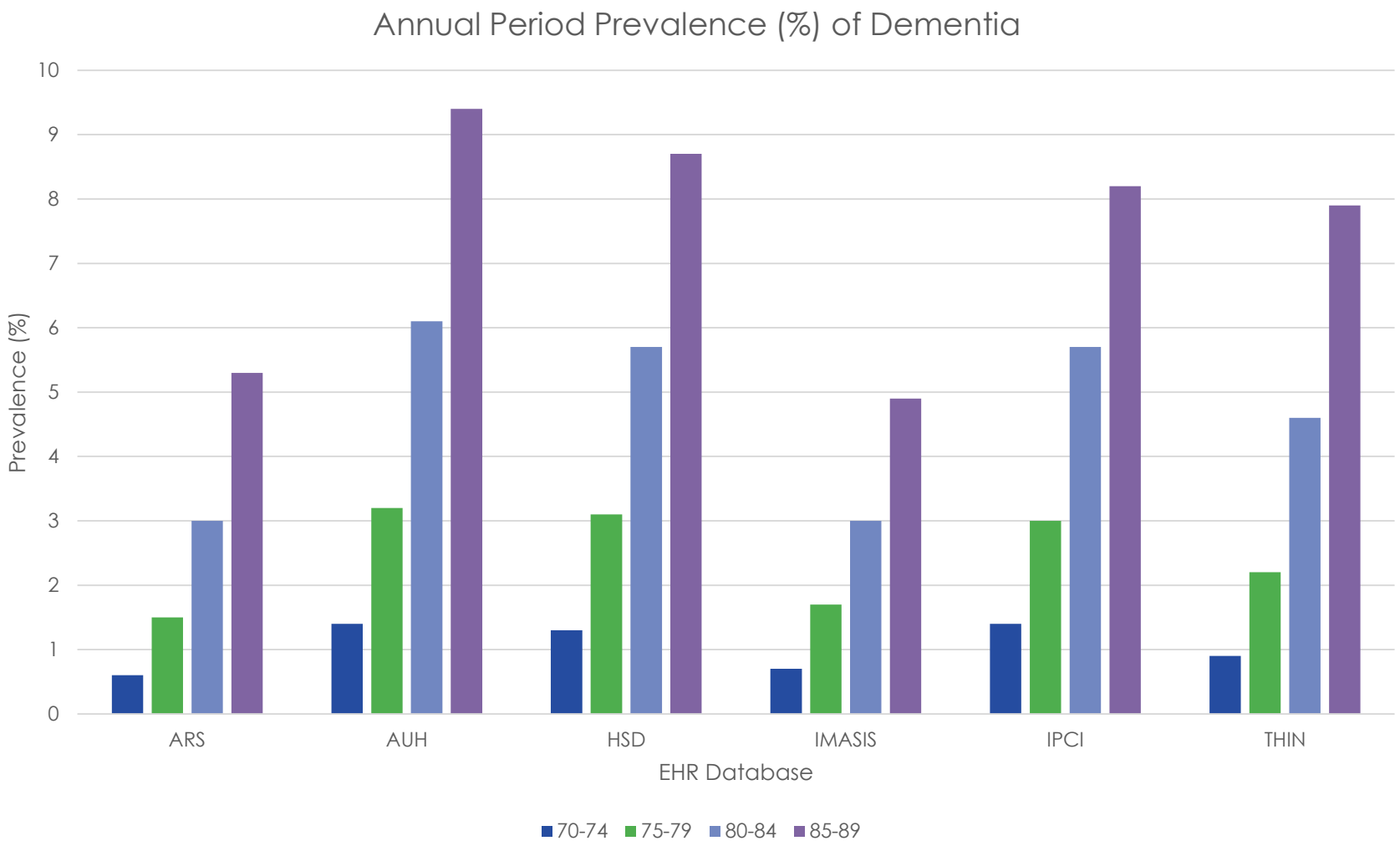


Incidence of Dementia

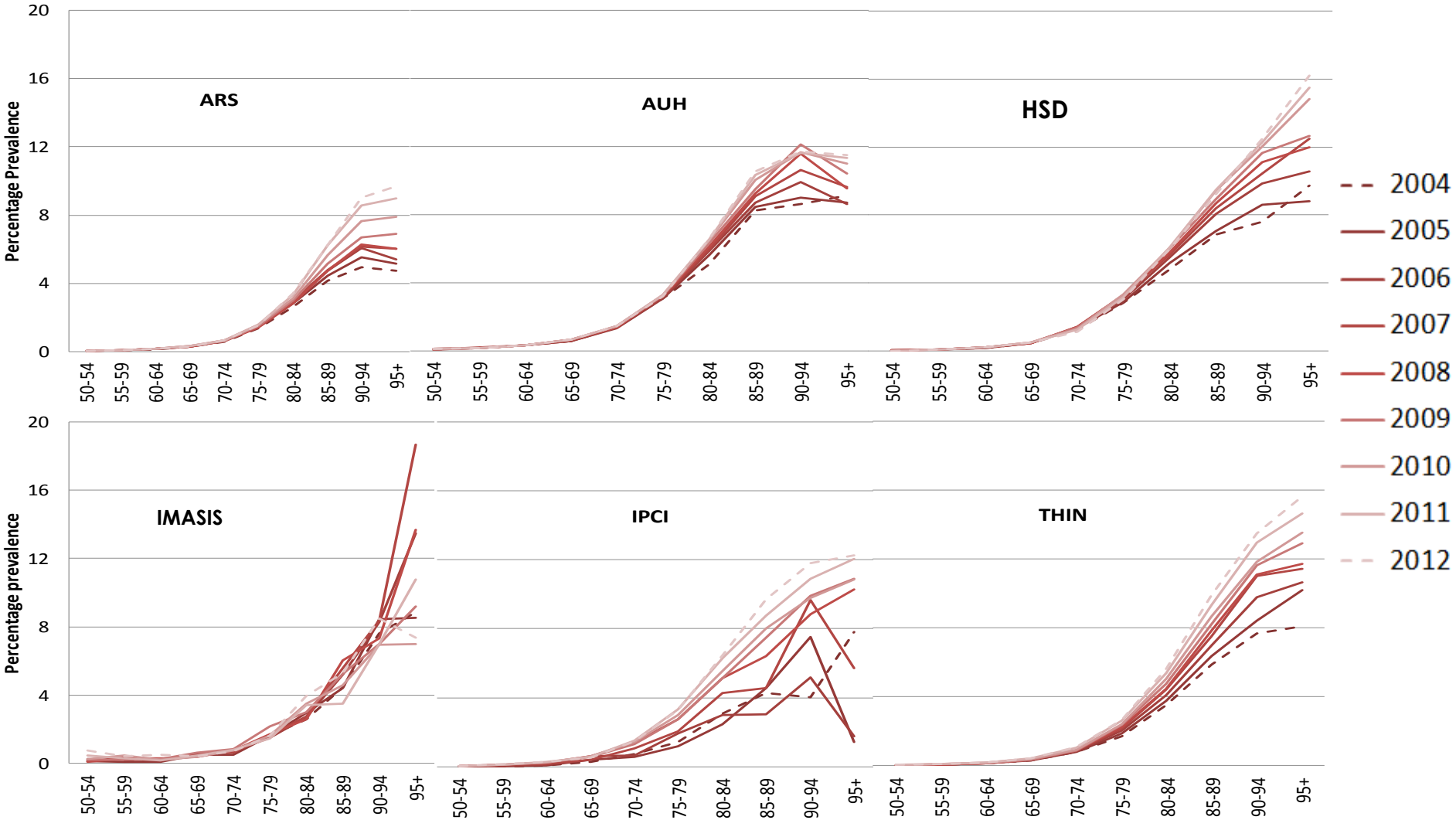


Annual incidence of first dementia diagnosis by age, year and EHR

Prevalence of Dementia



Prevalence of Dementia



One-year period prevalence of dementia by age, year and EHR

UC11: Levels of Blood Pressure, BMI and Total Serum Cholesterol Prior to Dementia Diagnosis

G Perera, U Gungabissoon, M Alexander, D Ansel, P Avillach, T Duarte Salles, MF Gordon, M Mayer, AJ Nevado-Holgado, GP Novak, A Pasqua, L Pedersen, A Ponjoan, P Rijnbeek, J Van Der Lei, R Stewart.

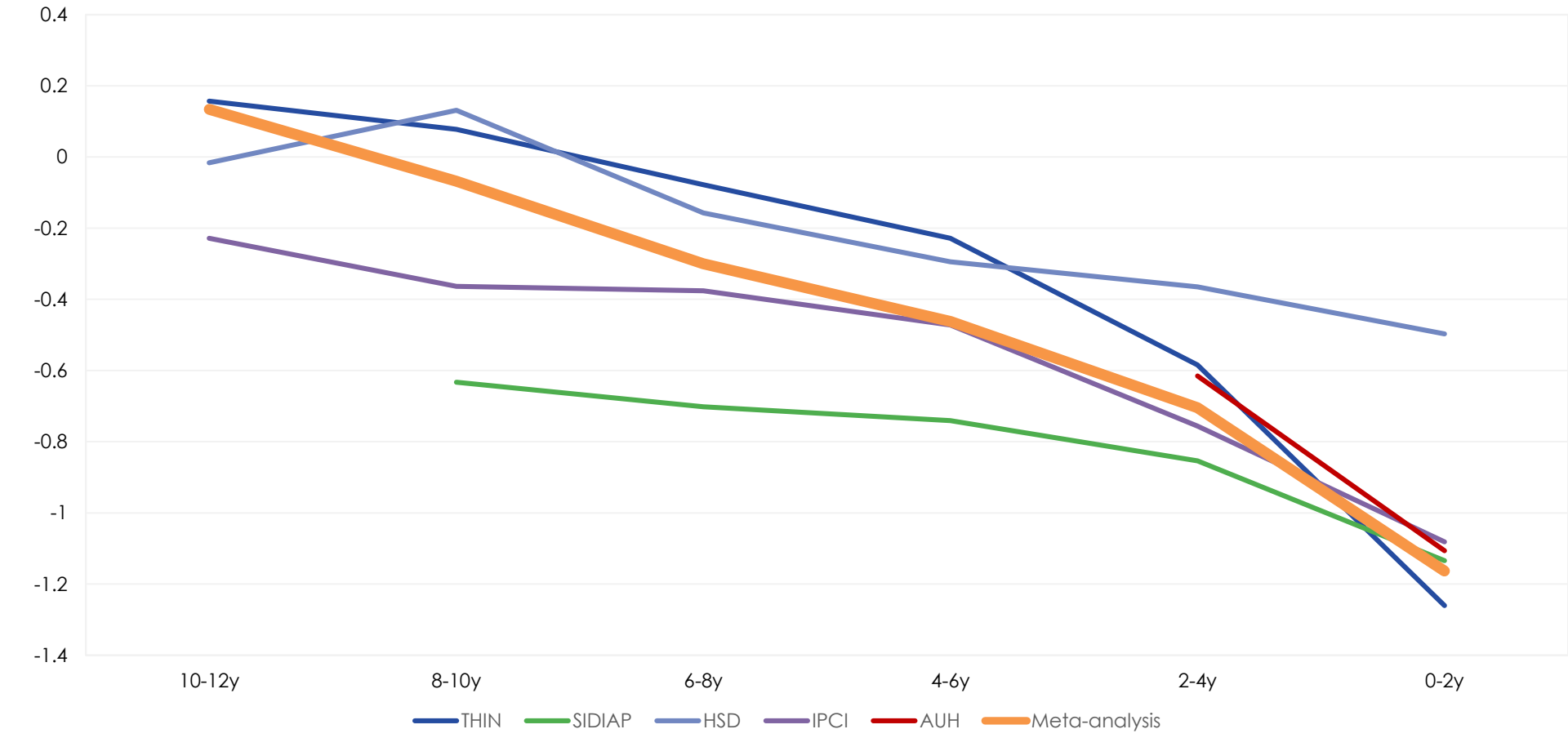
Poster presented at the AAIC in July in London

Background: Research cohorts have suggested changes in vascular risk factor levels prior to dementia onset – to be investigated in large-scale data sources.

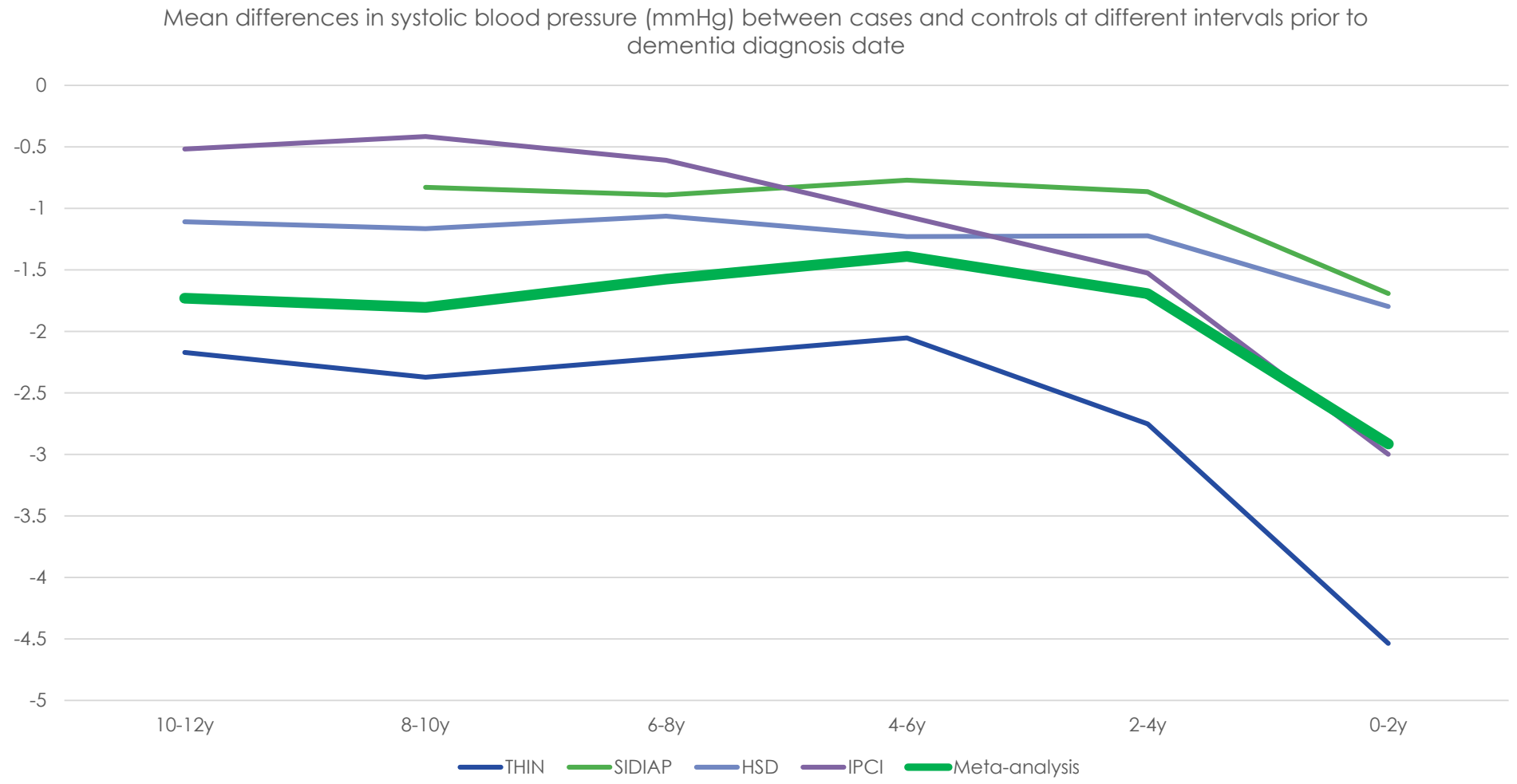
- 5 EHR databases analysed (AUH, IPCI, HSD, SIDIAP, THIN).
- An overall total of 287,000 cases of incident dementia compared to 28,700,000 age- and gender-matched controls on previously measured BMI, blood pressure, and total cholesterol.
- BMI and SBP show clear declines prior to dementia diagnosis – although with different patterns. DBP and total cholesterol are less consistent.

BMI Decline Prior to Dementia Diagnosis

Mean differences in BMI (kg/m2) between cases and controls at different intervals prior to dementia diagnosis date



SBP Decline Prior to Dementia Diagnosis



Summary and conclusion

- The wealth of data is impressive and the EMIF Platform provides a real opportunity for novel research
- Comparing results across data sources provides useful new insights and the basis for further research
- Platform tools developed so far work well and system integration is in the process of being tested
- Research efficiencies not realised yet as projects are being conducted during Platform development
- Useful experiences: Identified specific areas for improvement in study execution
- Sustainability of the Platform and its tools is the next goal


EMIF


efpia
imi
innovative
medicines
initiative

COFFEE BREAK

New Opportunities for Scaling Up Big Data Research

Chair: Prof Dipak Kalra

Institute for Innovation through Health Data

EMIF Strategic Data Extension Project - Key learnings

Tine Lewi & Omer Saka

Janssen Pharma R&D & Deloitte

European Medical Information Framework

EMIF Strategic Data Extension Initiative Key learnings

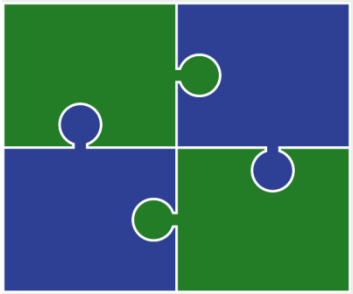
Realising the Value from Health Data ~ Improving Care and Research

September 21-22, 2017

Improving access to health data…
a key objective in EMIF Platform

Providing data access

- Scale
- Diversity
- Depth

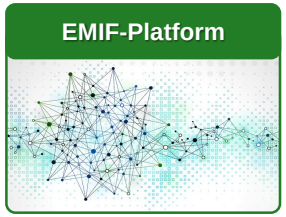


Delivering a working solution

- Privacy enabled solution
- Data harmonisation
- Analytical methods

Conducting relevant research

- Disease insights
- Value analysis
- Pharmaco-epidemiology



How to leverage EMIF methods and solutions potentially for other disease areas

Context

- EMIF demonstrates how to realize the value from health data in two key therapeutic areas: Alzheimer Disease and Metabolics.
- EMIF results are an illustration of how secondary re-use of human health data enables to address research questions which were previously very difficult to answer.

EMIF Achievements

Challenges

- EMIF Platform – Sustainability Workpackage - initiated in 2016 **a strategic data extension project** for further applying **EMIF Platform tools** (Catalogue, Workflows management, Harmonization to OMOP Common datamodel) and **governance framework** and address new research questions.
- A project has been carried out in collaboration with Deloitte to identify potential therapeutic areas being best candidates for EMIF strategic data extension program.
- This initiative aimed at identifying therapeutic areas with unmet needs, high potential for future collaboration being driven by common interests among EMIF members, research communities and data custodians, and promising application domains.

A step-wise approach was developed for identifying relevant disease areas, integrating views from distinct stakeholders

Screening of therapeutic areas/diseases

Screening disease area with rich pipelines, high unmet needs and areas of focus for research community and public sector.

1 **Pharma R&D Pipeline** 2 **Disease Burden** 3 **Public Health priorities research funding**

Assessing data availability

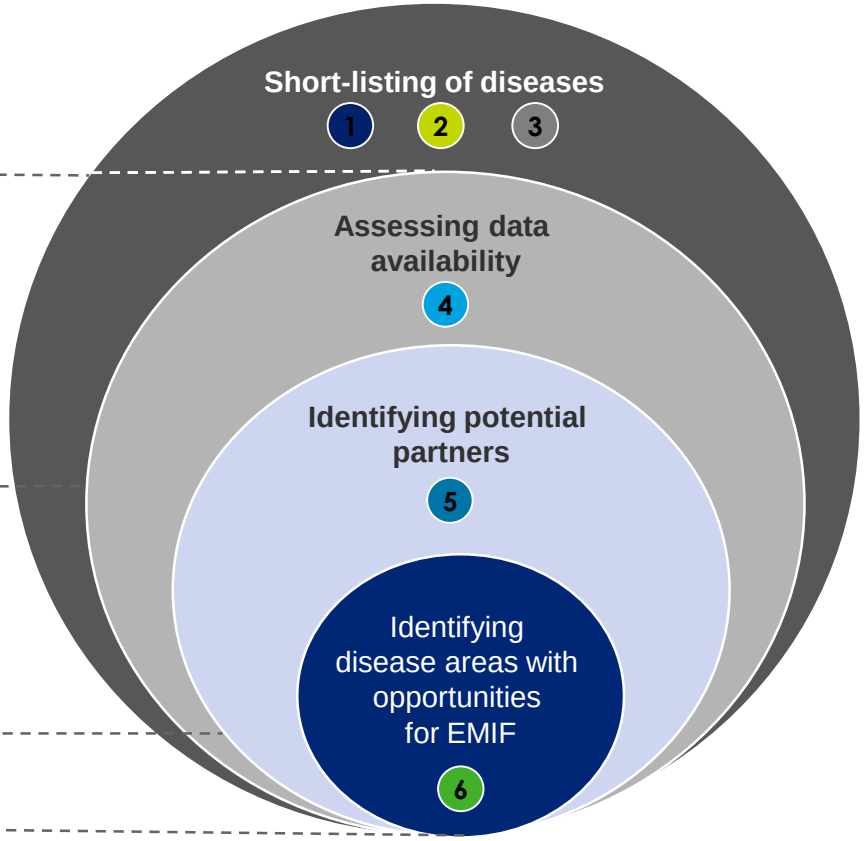
Identifying potential data sources to support secondary use of data per selected disease area by high-level screening and characterization of data sources.

Identifying research needs and potential partners

Gathering insights through 1:1 interviews with thought leaders on areas of research/application domains with great unmet needs and potential for collaboration.

Qualitative insights processing

Providing an evaluation framework to structure the insights and support internal alignment and assess potential opportunities.

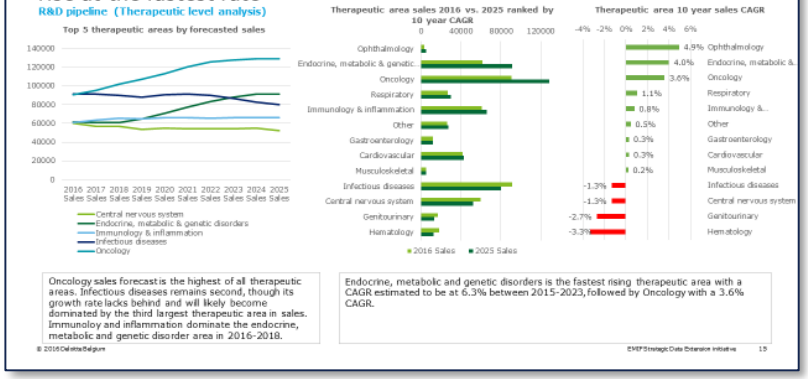


4 RWD available 5 Data partners 6 Impact assessment

Several aspects were gathered per high-level criteria and factored in as part of the short-listing process

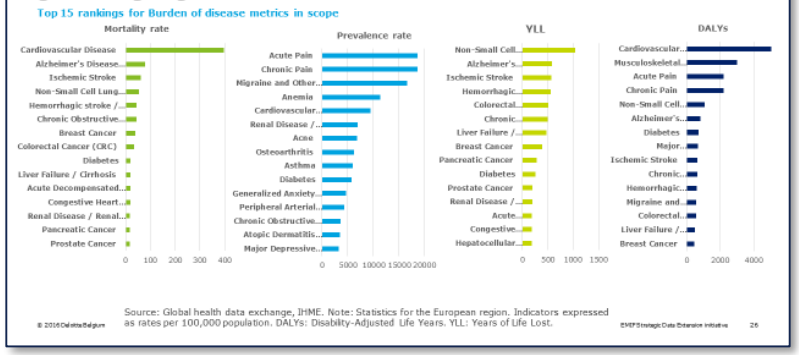
1 Pharma perspective

Assessing pipelines of Pharma – Oncology sales forecasts are the highest of all therapeutic areas, but Endocrine, metabolic expected to rise at the fastest rate



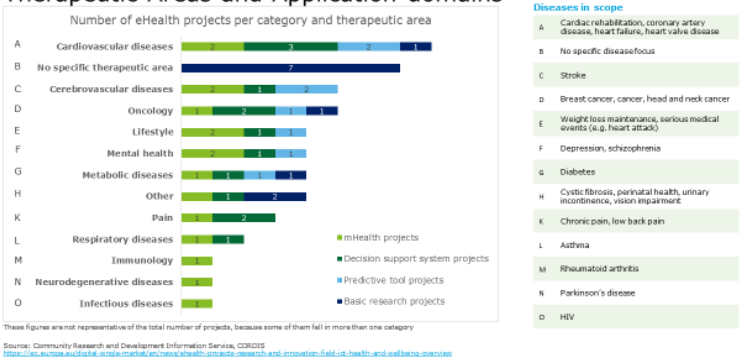
2 Burden of the disease

Cardiovascular disorders and Oncology as expected are associated with highest mortality rates in countries in scope gathering high number of DALYs.



3 Public Health priorities and research funding

eHealth European projects were mapped according to Therapeutic Areas and Application domains



National programmes (EU5) and European Joint Actions in specific therapeutic areas

Therapeutic areas						
Orphan diseases	✓ 2011-2016	✓ Launched in 2013	✓ 2013-16	✓ By 2020	✓ 2014-2018	✓ 2014-2019
Neurodegenerative diseases	✓ 2014-2019	✓ Final report in 2018	✓ 2014-2016	✓ 2012-2020	✓ Approved in 2016	✓ 2015-2017 (2nd phase)
Oncology	✓ 2014-2019	✓ Started in 2012		✓ 2015-2020		✓ 2014-2017
Cardiovascular diseases	-	-	✓ 2014-2016	✓ By 2020	-	-
Cerebrovascular diseases (Stroke)	-	-	-	-	-	-
Mental health	✓ 2013-2017	-	-	-	-	✓ 2013-2016
Infectious diseases	-	✓ By 2020	✓ By 2020	✓ By 2020	-	✓ Joint actions 2016
Pain (chronic, neuropathic)	-	-	-	-	✓ Approved in 2014	-
Respiratory diseases	-	-	✓	-	-	-
Hospital acquired infections (nosocomial)	-	-	-	✓ By 2020	✓ 2015-2020	✓ Joint actions 2016

Short-listed diseases were further refined based on strategic priorities

A set of scenario were derived to assess disease ranking per relevant perspective

	Sc. 1 All criteria		Sc. 2 Pharma pipeline only		Sc. 3 Pharma pipeline BoD		Sc. 4 BoD Public health /funding	
PHARMA PIPELINE	Inclusion	Weight (%)	Inclusion	Weight (%)	Inclusion	Weight (%)	Inclusion	Weight (%)
Pipeline: # Products in Ph II-III	yes	6%	yes	10%	yes	7%	no	0%
Pipeline: # Products in Ph I	yes	3%	yes	5%	yes	3%	no	0%
# Companies having Ph II-III products (EMIF)	yes	6%	yes	10%	yes	6%	no	0%
# Companies having Ph II-III products (Non-EMIF)	yes	3%	yes	5%	yes	3%	no	0%
# Companies having Ph I products (EMIF)	yes	6%	yes	10%	yes	6%	no	0%
# Companies having Ph I products (Non-EMIF)	yes	3%	yes	5%	yes	3%	no	0%
Therapeutic Areas Growth rate (2016-23)	yes	3%	yes	6%	yes	4%	no	0%
Therapeutic Areas Sales (2016)	yes	3%	yes	6%	yes	4%	no	0%
Therapeutic Areas Sales Forecasts (2025)	yes	3%	yes	6%	yes	4%	no	0%
Diseases Growth rate (2016-25)	yes	7%	yes	12%	yes	8%	no	0%
Diseases Sales (2016)	yes	7%	yes	12%	yes	8%	no	0%
Diseases Sales Forecasts (2025)	yes	7%	yes	12%	yes	8%	no	0%
BURDEN OF DISEASE (BoD)								
Deaths	yes	5%	no	0%	yes	7%	yes	15%
Prevalence	yes	10%	no	0%	yes	11%	yes	10%
DALYs (Disability-Adjusted Life Years)	yes	10%	no	0%	yes	11%	yes	10%
YLLs (Years of Life Lost)	yes	5%	no	0%	yes	7%	yes	15%
PUBLIC HEALTH								
EU- Joint Program	yes	2%	no	0%	no	0%	yes	10%
EU-5 National program (FR, IT, DE, UK, SP)	yes	1% / country	no	0%	no	0%	yes	3% / country
PUBLIC RESEARCH FUNDING								
EU- Research Funding Program	yes	1%	no	0%	no	0%	yes	10%
EU-5 Research Funding program (FR, IT, DE, UK, SP)	yes	1%	no	0%	no	0%	yes	3% / country
EU-eHealth Funding Program (disease-level)	yes	1%	no	0%	no	0%	yes	3%

A scenario analysis enabled ranking and short-listing TA/diseases candidates based on pipelines, BoD and priorities from public sector

	Scenario 1 (All criteria)		Scenario 2 (Pharma Pipeline)		Scenario 3 (Pharma Pipeline+BoD)		Scenario 4 (BoD+Public Health&Research)	
1	CLL /Small Cell Lymphocytic (SLL) - NHL	2.66	Non-Small Cell Lung Cancer (NSCLC)	3.00	CLL /Small Cell Lymphocytic (SLL) - NHL	3.00	Head and Neck Cancer	2.12
2	Breast Cancer	2.56	CLL /Small Cell Lymphocytic (SLL) - NHL	2.88	Dyslipidemia / Hypercholesterolemia	2.88	Gastric Cancer	2.09
3	Non-Small Cell Lung Cancer (NSCLC)	2.53	Prostate Cancer	2.88	Breast Cancer	2.88	Bone Cancer	2.09
4	Dyslipidemia / Hypercholesterolemia	2.49	Breast Cancer	2.88	Non-Small Cell Lung Cancer (NSCLC)	2.88	Uveal Melanoma	2.09
5	Prostate Cancer	2.46	HIV / AIDS	2.76	Diabetes, Type I	2.76	Cancer Pain	2.09
6	Diabetes, Type I	2.44	Dyslipidemia / Hypercholesterolemia	2.73	Prostate Cancer	2.73	Small Cell Lung Cancer (SCLC)	2.09
7	Colorectal Cancer (CRC)	2.43	Colorectal Cancer (CRC)	2.65	Colorectal Cancer (CRC)	2.65	Anal Cancer	2.09
8	Alzheimer's Disease (AD)	2.38	Ovarian Cancer	2.64	Gastric Cancer	2.64	Sarcoma	2.09
9	Gastric Cancer	2.37	Diabetes, Type I	2.64	Chronic Obstructive Pulmonary Disease (COPD)	2.64	Gastrointestinal Stromal Tumor (GIST)	2.09
10	Major Depressive Disorder (MDD)	2.33	Multiple Sclerosis (MS)	2.52	Alzheimer's Disease (AD)	2.52	Neuroendocrine Tumors (NET)	2.09
11	Chronic Obstructive Pulmonary Disease (COPD)	2.32	Hepatitis C (HCV)	2.52	Major Depressive Disorder (MDD)	2.52	CLL /Small Cell Lymphocytic (SLL) - NHL	2.09
12	Indolent Non-Hodgkin's Lymphoma - NHL	2.21	Asthma	2.46	Hemophilia A	2.46	Merkel Cell Carcinoma	2.09
13	Asthma	2.16	Renal Cell Cancer (RCC)	2.41	Indolent Non-Hodgkin's Lymphoma - NHL	2.41	Marginal Zone Lymphoma - NHL	2.09
14	Ovarian Cancer	2.12	Chronic Obstructive Pulmonary Disease (COPD)	2.41	Myelofibrosis (MF)	2.41	Mantle Cell Lymphoma - NHL	2.09
15	Small Cell Lung Cancer (SCLC)	2.10	Alzheimer's Disease (AD)	2.40	Asthma	2.40	Breast Cancer	1.99
16	Hemophilia A	2.10	Gastric Cancer	2.39	Ovarian Cancer	2.39	Colorectal Cancer (CRC)	1.99
17	Head and Neck Cancer	2.08	Multiple Myeloma (MM)	2.34	Respiratory Syncytial Virus (RSV)	2.34	Benign Prostatic Hyperplasia (BPH)	1.99
18	Respiratory Syncytial Virus (RSV)	2.06	Major Depressive Disorder (MDD)	2.29	Small Cell Lung Cancer (SCLC)	2.29	Clostridium difficile Infection (CDAD/CDI)	1.98
19	Myelofibrosis (MF)	2.06	Schizophrenia	2.28	Fabry's Disease	2.28	Respiratory Tract Bacterial Infections (Excluding Pneumonia)	1.98
20	Acute Coronary Syndrome (ACS)	2.01	Hodgkin's Lymphoma	2.13	Osteoarthritis	2.13	Bone and Joint Bacterial Infections	1.98

Notes: COPD: CLL/SLL: Chronic Lymphocytic Leukemia (CLL)/ SLL (SLL) – NHL: Chronic Lymphocytic Leukemia (CLL)/Small Cell Lymphocytic Lymphoma (SLL) – NHL

*Short-listed diseases (not exhaustive)**

Immunology & inflammation

- Osteoarthritis
- Multiple sclerosis

Endocrine, metabolic

- Non-Alcoholic Steato-hepatitis
- Osteoporosis

Oncology (solid tumors)

- Breast cancer
- NSCLC
- Head & neck cancer
- Gastrointestinal Stromal Tumor (GIST)

Hemato-oncology

- CLL/Small Cell Lymphocytic Lymphoma (SLL) - NHL
- Mantle Cell Lymphoma

Mental health

- Major Depressive Disorder (MDD)

(*): listed unranked

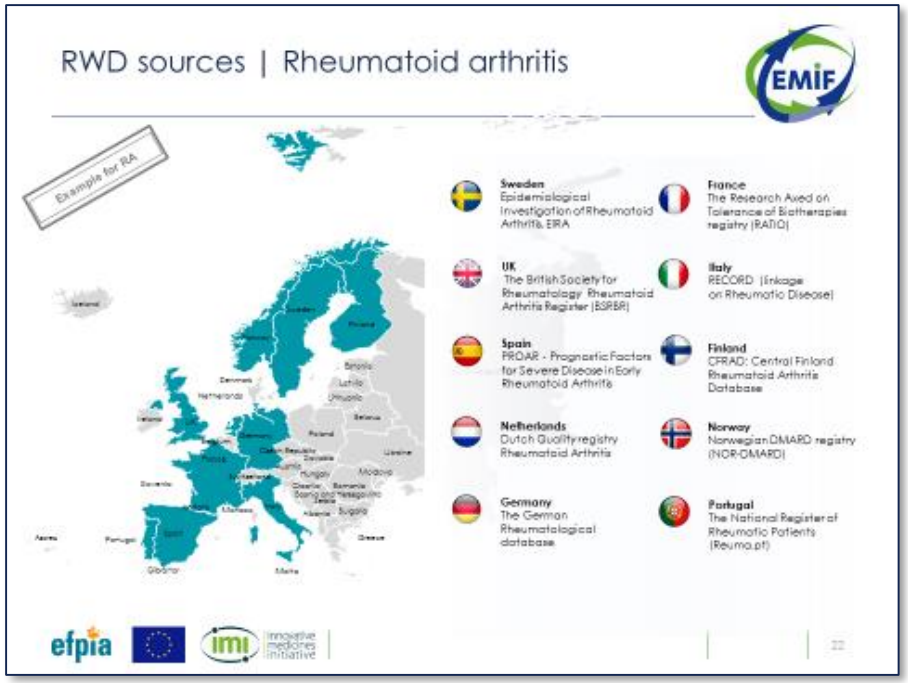
RWE data availability was assessed for the short-listed diseases

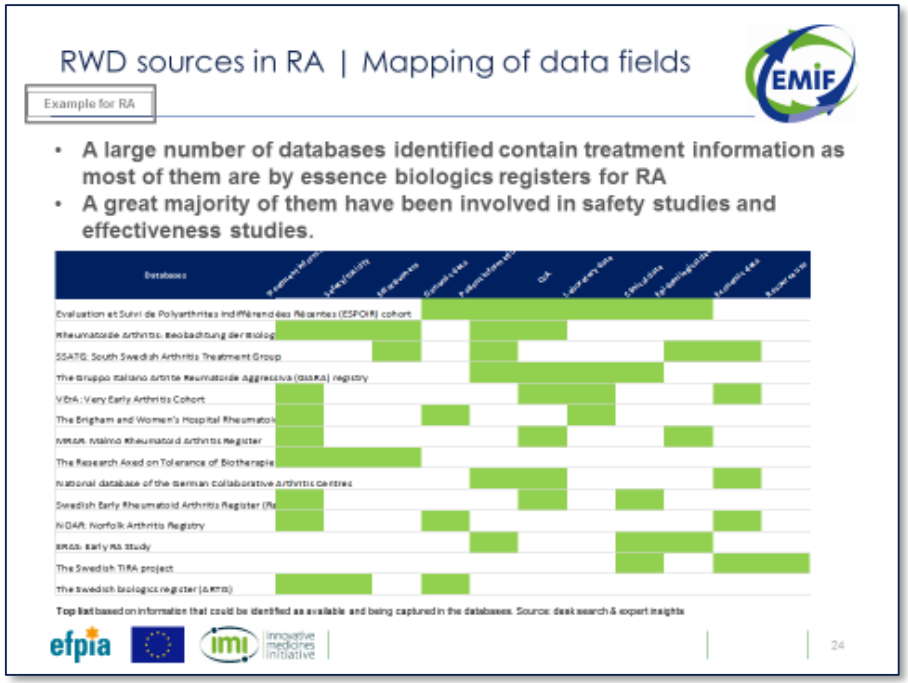
1 Step 1: Providing data source synopis

- Data owners
- Volume of data: #Patients
- Investigator(s) contact details
- Depth of data fields: (epidemiology, clinical, treatment information,…)

2 Step 2: Mapping data fields, access conditions

- Additional data collected (synopsis)
- Year of establishment
- Type of outcomes
- Feasibility to retrieve condition(s) of access and previous collaboration with Pharma





Interviews with disease experts shed light on challenges and research domains with high unmet needs

1:1 interviews conducted in three disease areas and relevant domain expertise…

Experts with clinical expertise

- Rheumatoid arthritis
- Multiple sclerosis
- Oncology

Experts with domain expertise

- Biologics registries
- Population-registries
- Biobanking

…brought qualitative insights at several levels

- Inspiring data-sharing initiatives
- Research question(s) / Application domains* with need of increased collaboration
- Past collaborative studies achievements
- Potential relevant stakeholders for partnerships
- Challenges related to collaboration and secondary use of data

(*): application domains: biomarker, Comparative effectiveness, Safety monitoring, Disease understanding, Population segmentation

Learnings from the experts interviews

Experts particularly highlighted the following points…

Consensus on the need of Catalogues of real world datasets

- There is **value of establishing catalogues/ inventories** to report datasets characteristics, data-sharing and partnership rules.
- There is a **need to strengthen researchers and data custodians networking** and data suitability assessment.
- Absence of IT-infrastructure is not a limiting factor by all experts interviewed.

Data Quality is a key factor not fully addressed yet

- **Quality management should be ensured,** with further effort warrranted at several levels (i.e. patient examination, biological sample level, ab-test, imaging).
- FAIR principles (Findable, Accessible, Inter-operable, Re-usable) should incoporate a **Quality pillar**.

IT-related issues are less problematic than data governance and legal hurdles

- Importance of promoting a framework that safeguards local data control.
- Risks of loosing data ownership and uncontrolled data dissemination were also highlighted as matter of concerns.
- Personal data protection generates uncertainty in the perspective of the rolling out of the new EU GDPR.

A need of further collaboration supported by enhanced data sharing/re-use in specific research domains

Multiple sclerosis…

- **Long-term safety studies**
- **Predictive analytics and patient stratifications**
- **MS and pregnancy**

- Safety issues likely one of the **application domains with highest common need** to collaborate.
- Need for further identifying of**predictive factors of disease onset, progression** and **treatment response**.
- Ongoing MultipleMS and BigMS data collaborative studies.

Rheumatoid arthritis…

- **Long-term safety studies**
- **Disease phenotyping Patient segmentation**
- **Biosimilars**

- Safety: **even large registries** likely willing to collaborate in to monitor **(very) rare adverse events**.
- Investigating **predictors of disease progression** and **treatment response will imply data linkage with biobanks and enriched data sets**
- In the field of BioS, understanding switching patterns and addressing safety issues.

Oncology…

- **Enhanced epidemiology surveillance**
- **Monitoring implementation of clinical guidelines**
- **Monitoring use of cancer drugs in real life**

- Key potential areas of focus beyond disease epidemiology.
- Data harmonization is needed for tumour characterisation, staging and summary treatment.
- Enhanced data linkage between biobanks and oncology population-based registries.

Challenges were also highlighted by experts, with some being reported for both diseases

Challenges across diseases

Inventory/Catalogue with detailed are still lacking of cancer registries with detailed synopsis is lacking

- Still lacking even in the field of oncology population-based registries.
- Data-linkage, conditions of access and partnerships shall be described in depth.
- A catalogue set up for RA (2006-07) but keeping the information up-to-date in the long run is challenging

All fields would benefit from enhanced standardization

- Enhanced standardization in the way data are collected (e.g. exposure being measured and reported).
- A critical preliminary step for the future is to build consensus on standard clinical assessment tools to increase consistency of data collection (exposure, patient outcomes).

Varying degree of collaboration with private sector and several conditions for successful collaboration to be met

- Data governance represent a major hurdle beyond technical issues.
- Scientific award and co-authorship increasing scientific reputation of study participants and ability to receive research grants and sponsorships.
- Keeping control on the data to mitigate the risk of loss of funding if the governance of the data is not guaranteed

A federated network of databases has been pointed out as a possible solution to address needs of further collaboration

- Data governance represent a major hurdle beyond technical issues.
- Willingness to collaborate will depend on the inability to ensure sustainable funding, fair governance framework with no loss of control on the data with mandatory delegation to a third party

The insights from the research were ultimately structured into a heat-map

Sources of information	1. Insights from EFPIA members	2. Real world data available	3. Experts insights completed by survey

HEAT-MAP	RA	MS	Oncology	IBD	Liver diseases
Alignment with EFPIA members	high	medium	high	high	high
Alignment on focused research questions	high	medium	low	medium	(na)
#datasets identified	high	high	high	high	high
Experts interviews response rate	medium	high	medium	no response	(na)
Need for an inventory in the field	high	high	high	high	(na)
Need for networking	high	medium	high	high	(na)
Need for an IT-infrastructure for data-sharing and re-use	medium	medium	medium	(na)	(na)
Existing collaborations with data-sharing	high	high	high	medium	(na)

Distinct color-grading (indications due to number of responses in IBD and oncology (n=7 respondents each).

Several drivers for enhanced collaboration within the scope of EMIF have been identified

Drivers for success and opportunities

Rheumatoid arthritis…	Oncology
A **strong alignment within EFPIA members** with members likely willing to steer next steps.	Strong **alignment within EFPIA members with leardership** comittee foreseen within EMIF.
Relevant data sources (**biologics registries**) were identified with precedent in **collaborative studies** implying data-sharing in the field of **safety and effectiveness**.	In absence of publicly available catalogue, deep diving into all cancer registries and cancer specific research data sources is a very resource-intensive task.
Existing **EULAR Taskforce on Biologics Registers** to enhance collaboration: share best practices, maximize quality, and provide infrastructure to enable methods development.	**First contact initiated with chairman of the European Network of Cancer Registries,** willing to build follow-on discussion with EMIF.
No existing large-scale collaborative study funded yet by EU (e.g. Horizon 2020) or likely planned in the near future.	A new Health Information System planned to be launched in 2017 by ENCR as a new platform to disseminate relevant information to the research community and general public.

Biosimilars are an area of opportunity for research

Building a cross-country collaborative study of biosimilars in Rheumatoid arthritis and IBD:

Treatment patterns – Safety – Effectiveness

*1. **High attractiveness** with likely **endorsement by multiple stakeholders** given implications of providing real-world evidence based assessment in the field of biosimilars to foster successful adoption of most appropriate treatments by mitigating patient acceptance and building prescribers confidence (in collaboration with payers)*

*2. **A cross-application domains** topic with feasibility to move towards **Predictive analytics***

*3. **Feasibility to expand beyond RA** to **other rheumatic diseases** to broaden the reach of the research and maximize impact.*

In this changing landscape EMIF and further programs have a significant role to play

1 Maturity of private, curated data through large-scale investments

- HLI gains $220 million investment through investors including Illumina, Celgene and GE Ventures
- PatientsLikeMe secured a $100+ million investment through partnership with iCarbonX
- Merck, has joined the Oncology Research Information Exchange Network® (ORIEN)
- Verily – Alphabet baseline, BGI/BC

2 Governments are investing in generating patient data sets

- Qatar is establishing a genome map of the local population (6000 genomes sequenced by mid 2017)
- China confirmed precision medicine to be part of its Five Year Plan for 2016-2020
- 21st Century Cures – 1M pt. cohort+, IMI, UK 1000 genomes
- Small geographies: The Human Project is studying the lives of 4,000 NYC households over the span of decades across domains

3 HIT companies looking to monetize data

- Cerner Healtheintent – collects data from disparate sources for pop health and precision medicine
- Phillips using AWS to analyze and store 15 PB of patient data gathered from 390 million imaging studies, medical records, and patient inputs
- Higi – kiosks in pharmacies that tracks trends and changes in body data available in partnership with pharmacies

Drivers

- Land grab and competitive move by private ventures and nations to achieve market ownership of channels in HCLS consuming key data sets and obtaining critical mass of data
- Costs lowering of high content testing creating
- Businesses with infrastructure today must seek subsidy in a world where capital/infrastructure costs are penalized vs. pure virtualized cloud operations
- Large healthcare spend entices data investments

Considerations

- Expect any software business with a cloud model to have a data approach. Assume it is available or coming.
- Disposition to private data aggregators and federal initiatives are 'big bets' – opportunity from being a scale buyer to drive agenda
- Platform, tools, and talent to make full use of data


EMIF


efpia
imi
innovative
medicines
initiative

Patient/Citizen Generated Health Data: The Next Real World Data Frontier

Alison Bourke

QuintilesIMS

Patient/Citizen Generated Health Data: The Next Real World Data Frontier

Alison Bourke

Scientific Director, RWI, QuintilesIMS

Madrid, 22 September 2017

Why now?

- Digital social communication opens new channels to/from patients/citizens
- Factors outside the formal clinical environment (eg social deprivation, exercise, diet) have a huge health impact and E/M-health facilitates access to more routine and granular data outside of clinical setting eg Fitbit
- Patientcentric view

Patient Generated Data

Much RWD is EMR, and some limitations of EMR alone:-

- May not be dispensed prescriptions so compliance unknown
- Limited information on OTC medications
- Limited data on non-routine care, lifestyles, diet
- Limited information on how patients feel
- No information on patients' health/life priorities
- Little data on environment eg climate, pollution

Summary – Snapshot data from the Healthcare team view

Patient/Citizen Generated Data
(Also known as Patient Generated Health Data - PGHD)

Social media

Apps

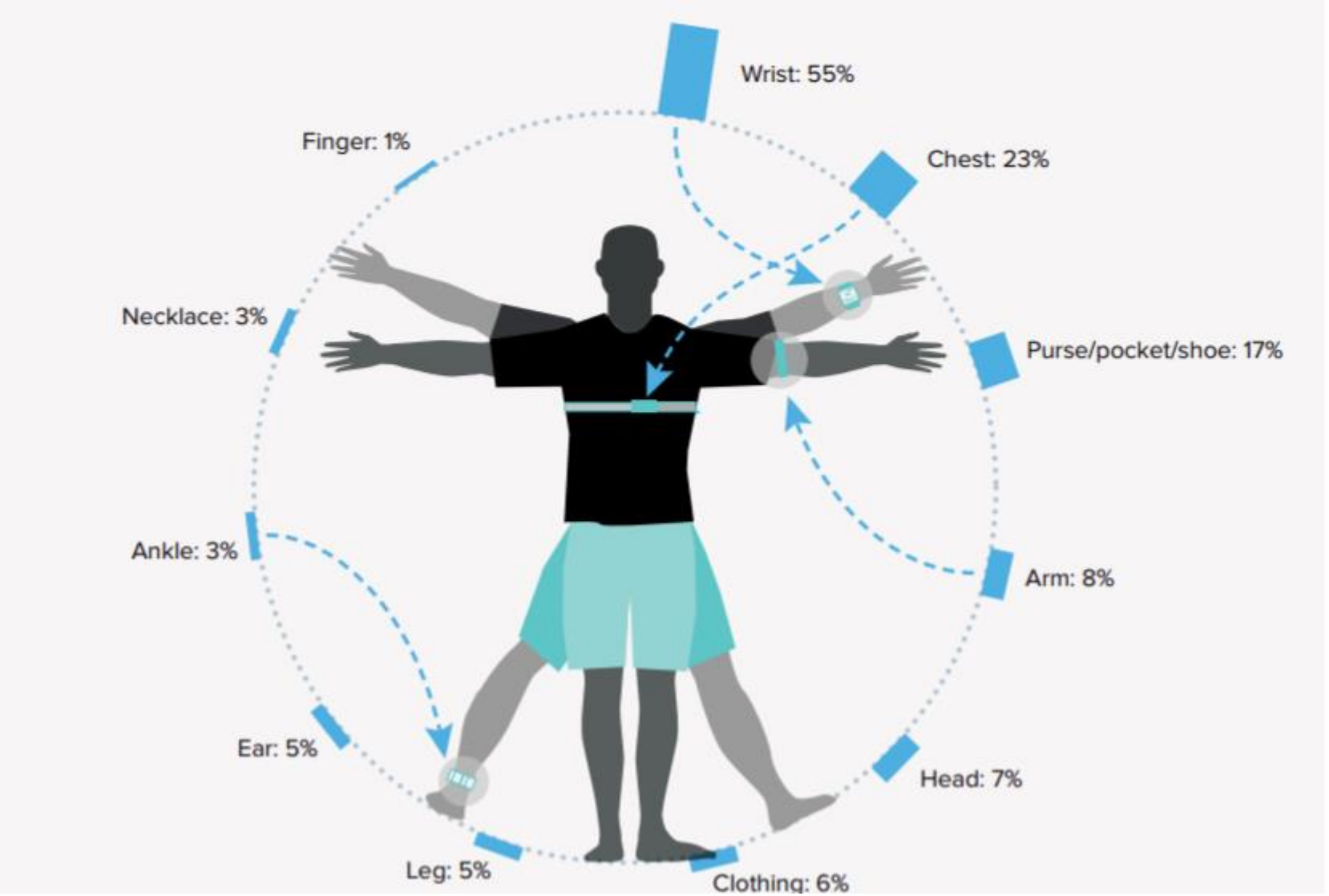
Bio-sensors

IOT

Environmental sensors

Wearables

Location of Wearables



Types of Patient Generated Data

- Biometrics outside of clinical setting
- Environmental factors

  Example: MyAirCoach- the use of home-monitoring and mHealth systems to predict deterioration in asthma control and the occurrence of asthma exacerbations: Honkoop et al *BMJ* 2017

- Lifestyle - fitness, diet, sleep

  > 160 Fitbit ClinicalTrials.gov studies – including obesity, cancer, post surgery

- Adherence & compliance of treatment
- Qualitative data - QOL & values - PROMs

  Example: Cloudy with a Chance of Pain

- Volume
- Variety
- Velocity

Issues/Challenges of PGHD (1 of 2)

- Cost
- Sample Bias
- Patient Recruitment
- Patient Retention
- Data Access
  - Technology
  - Ownership
  - Consent models
- Confidentiality
- Identifiability
- Device Standardisation

Issues/Challenges of PGHD (2 of 2)

- Keeping up with fast development
- Data Standardisation
- Data Quality
  - Subjectivity
  - Completeness
  - Accuracy
- Cyber security
- Workflow
- Analytics, eg NLP, ML, data visualisation
- Liability of actionable insights.

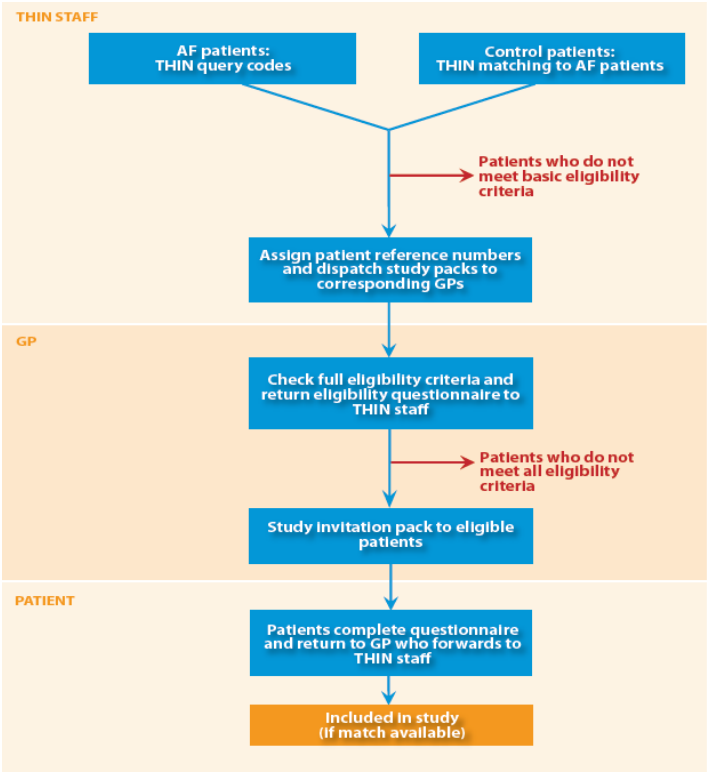
Direct Patient data + EMR: Example 1 - AFLOAT

Atrial Fibrillation Longitudinal Outcomes Assessment Study

To assess the symptom burden of AF in newly diagnosed patients identified within one week of symptom recording

516 case – control pairs were identified as soon as data was received from THIN practices

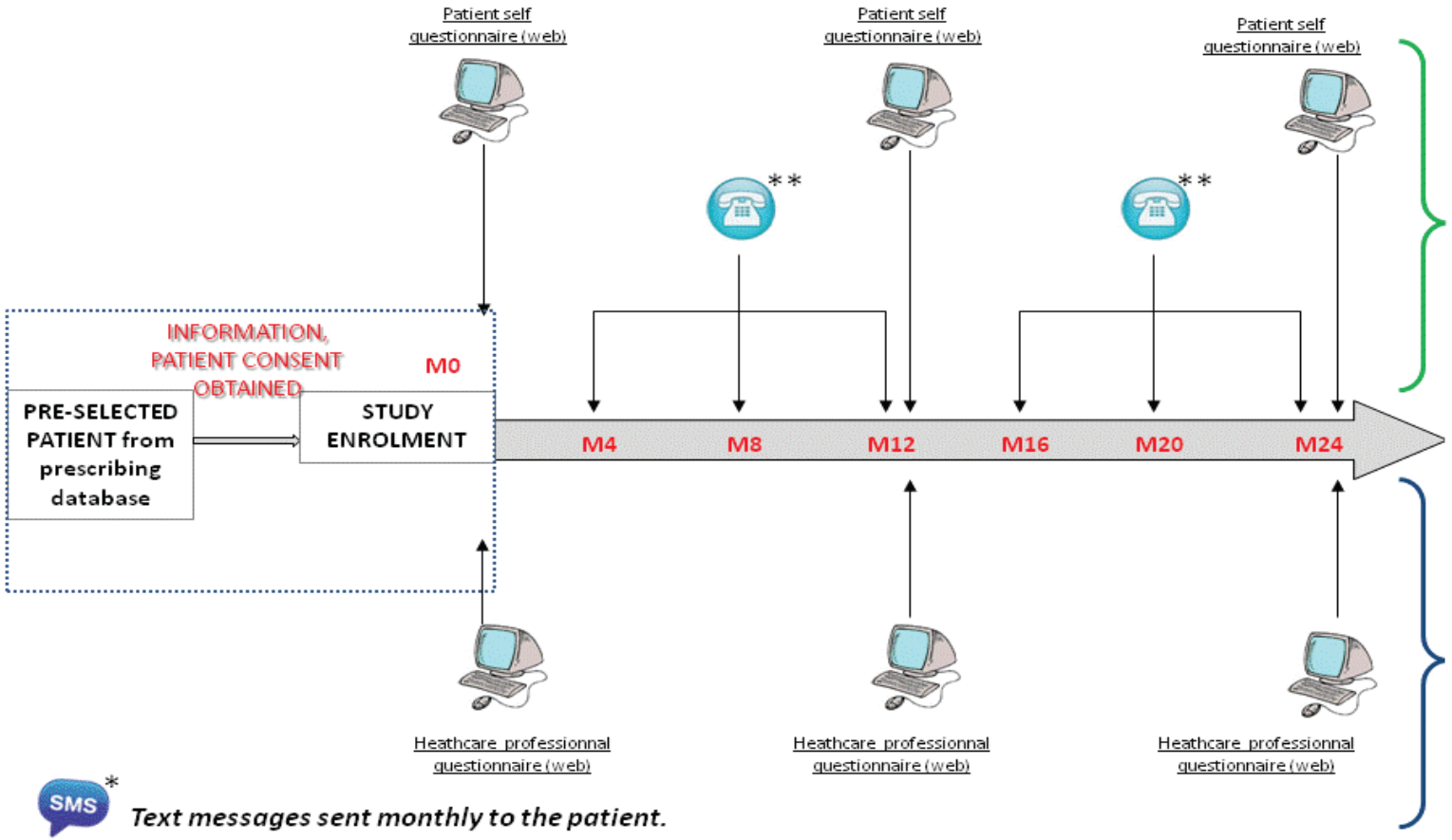
→ 82% GP response rate and 50% patient response rate to questionnaires forwarded



Vial D, Thompson M, Cockbain C, Hogan S, Johnson M, Bourke A, 2013. **Rapid identification and recruitment of patients from The Health Improvement Network (THIN) primary care patient data for a health-related quality of life (HRQOL) study of patients with atrial fibrillation (AF) in the United Kingdom.** *Value in Health* 16(3), A45-A46.

Direct Patient data + EMR: Example 2 - ASTRO-LAB

Assessment of the safety of LABAs in asthma in routine care by combining health-care databases and direct patient follow-up



124

Direct Patient data + EMR: Example 3 - PROTECT

An exploratory study of self-reported medication use in pregnant women and pregnancy outcomes with validation of self-reported data through electronic health records & national prescription data

- Recruitment via leaflets in pharmacies, pregnancy websites, advertising
- Internet v phone
- Self-reported medication use (including non prescription eg herbal, illicit) compared with data from electronic health records, national prescription data, and regional prescribing practices

The PROTECT project received support from the Innovative Medicines Initiative Joint Undertaking (www.imi.europa.eu)

Direct Digital Patient Involvement

Offers fantastic research and clinical care advantages:-

- Improved recruitment/retention to research
- Supplement traditional RCT and EMR data
- Increase the frequency and accuracy of data capture
- Assist early detection and diagnosis
- Inform clinical pathways, drug usage and utilisation
- Support precision medicine – targeted therapy
- Improve & inform medicines adherence
- Prioritise care in line with patient view

New projects….?

Crossing the new RWD frontier

More than just new data…..

Improved research & clinical care

Data fuelled apps managing disease

New carer – patient dynamic

Any questions?

Thank you

alison.bourke@quintilesims.com


EMIF


efpia
imi
innovative
medicines
initiative

Ethical Considerations Within Federated Data Use

Prof Dipak Kalra

Institute for Innovation through Health Data

The trustworthy scaling of big data research

Dipak Kalra
Bart Vannieuwenhuyse, Janet Addison, Nige Hughes, Caroline Sage,
Nathan Lea, Louis Schilders, Kathleen Fadden

efpia innovative medicines initiative

If we are to scale up big health data research, across data sources, across countries

- Trust is needed to protect the interests of
  - Data subjects
  - Data sources
  - Research users
  - Society as a whole

Components enabling the trustworthy reuse of health data for research

- Bona fide (societally acceptable) purposes
- Bone fide research organisations
- Transparently defining the source data: FAIR principles
- Precisely specifying the intended research
- Complying with research ethics and consent
- Protecting the identity of data subjects
- Agreeing terms for recognition and reward
- Compliance and audit
- A social contract?

The EMIF Code of Practice (ECoP)

- Developed in order to help ensure:
  - that the EMIF Platform and Services are used in ways that comply with legislation and policies on data protection
  - that EMIF upholds best practices in the protection of personal privacy and information governance
  - that EMIF promotes best practices in the conduct of clinical research using health data, for the public good

- We expect to contribute this into a wider European governance landscape for research using big health data

Bona fide research

- The key characteristic of bona fide research is that its objective is to discover new knowledge intended for the public good and to be made publicly accessible (i.e. published)
- A bona fide research organisation is one that is appointed or accredited or funded to undertake bona fide research, and/or has made public its commitment to adhere to recognised research governance principles.
- It is not a requirement that such research is the primary business of that organisation, or that all of the research undertaken by that organisation is published. It is not a requirement that the organisation be publicly funded.
- New knowledge includes the corroboration of, or the challenge to, existing knowledge as well as completely new discoveries
  - intermediary stages of the research life cycle might not be made publicly accessible

EMIF research users seeking health data access will be verified to be members of bona fide research organisations who have legitimate purpose in conducting research queries on health data

The EMIF Charter - principles shaping the ECoP

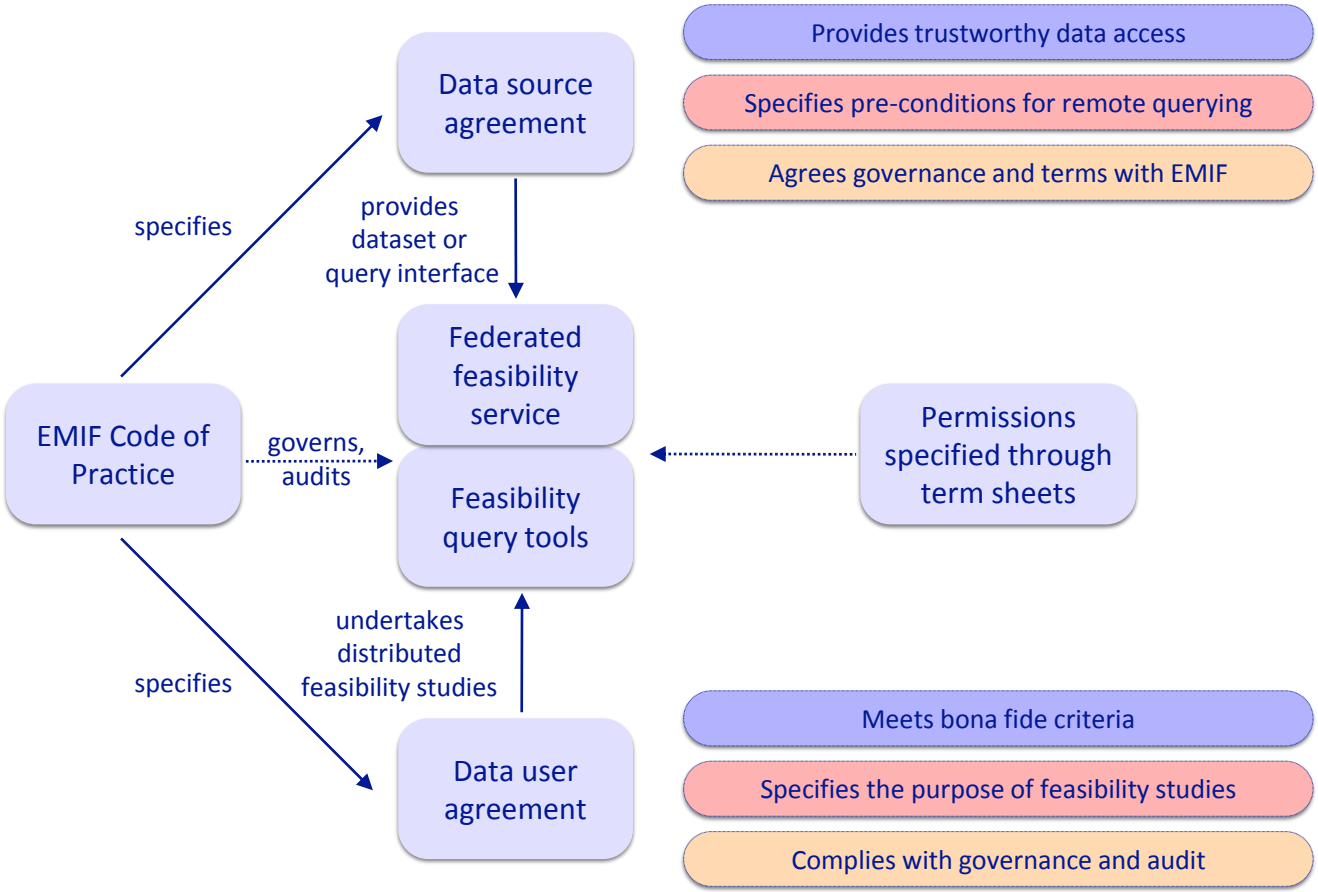
- The EMIF platform can only be used, for assessing the feasibility of a study and for conducting research, by bona fide research organisations and for the objective of discovering new knowledge intended for the public good and to be made publicly accessible (i.e. published)
- Data sources
  - will always have autonomy over which data are made accessible and for which types of research
  - will always determine ethical acceptability and scientific validity
  - must be transparent about their data
- Data users
  - must adhere to the ethical rules and privacy protection policies of each data source
  - may only use the data for the specific agreed research purposes
  - must acknowledge the sources of the data they have used, and EMIF

The FAIR Guiding Principles

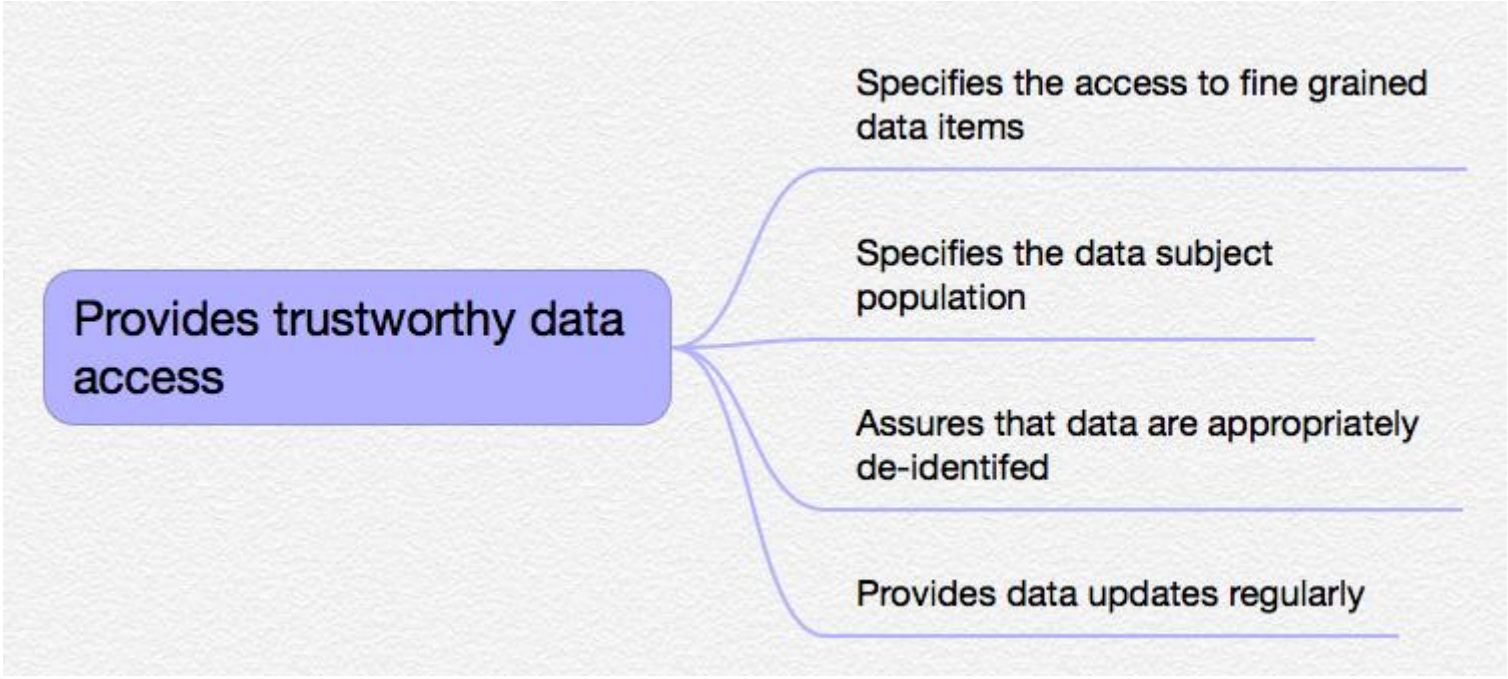
- To be Findable
- To be Accessible
- To be Interoperable
- To be Reusable

- can be applied to access to health data, for research

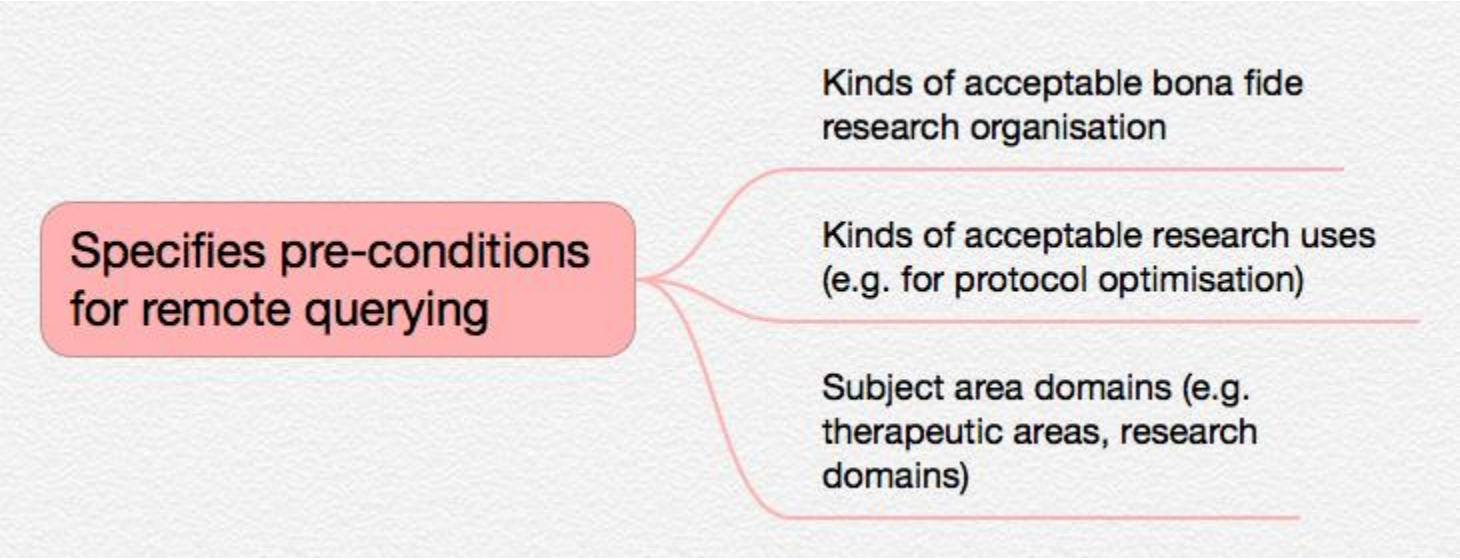
Governing EMIF Feasibility Services



Feasibility: data source obligations



Feasibility: data source obligations



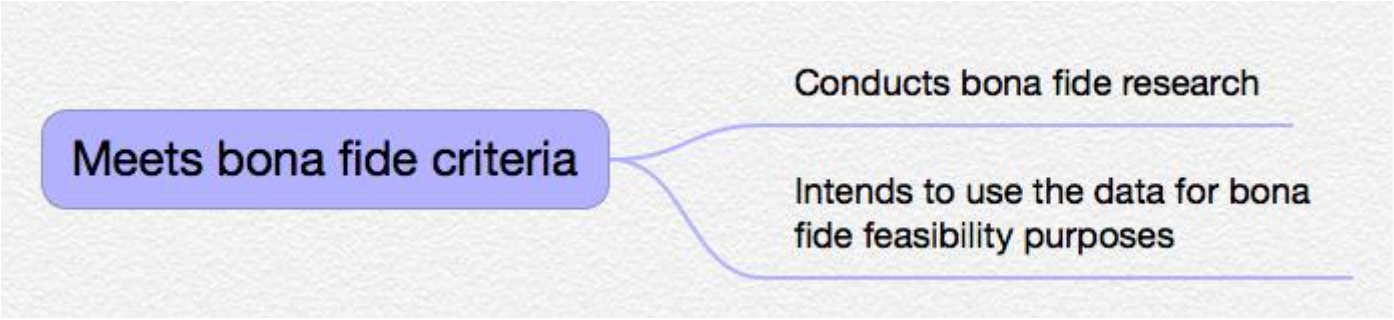
Term list for specifying kinds of research organisation

- Pharma company
- Medical device manufacturer
- ICT company
- Regulatory body
- Academic research organisation
- Payer
- Government department
- Patient associations and charities

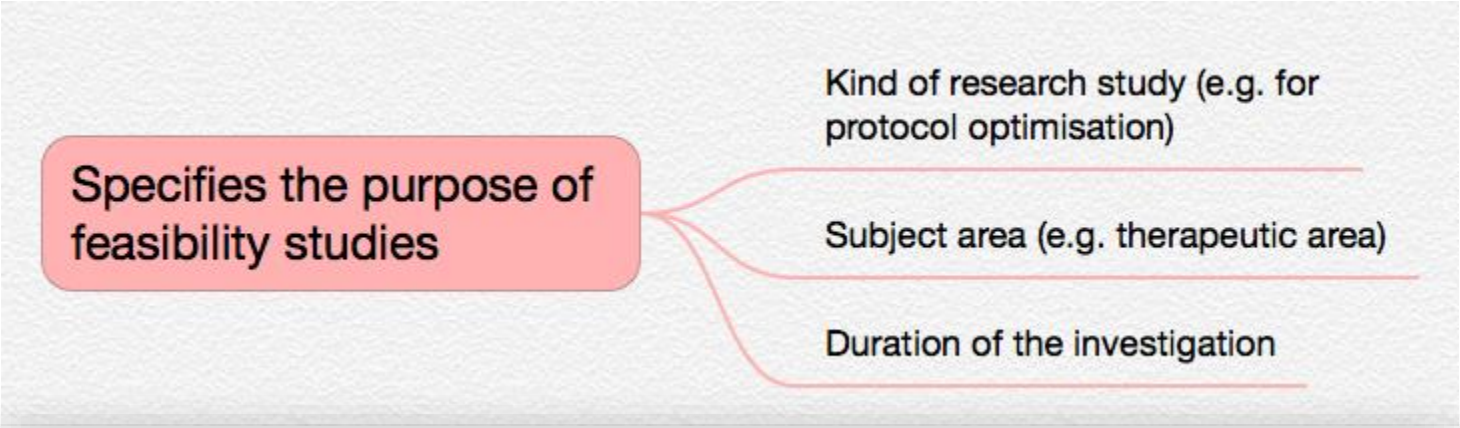
Term list for specifying types of research study

- Observational/non-interventional
- Interventional
- Comparative effectiveness
- Health economic studies
- Market research
- Post-authorization Safety Studies
- Post-authorization Efficacy Studies
- Pharmacogivilance

Feasibility: data user obligations

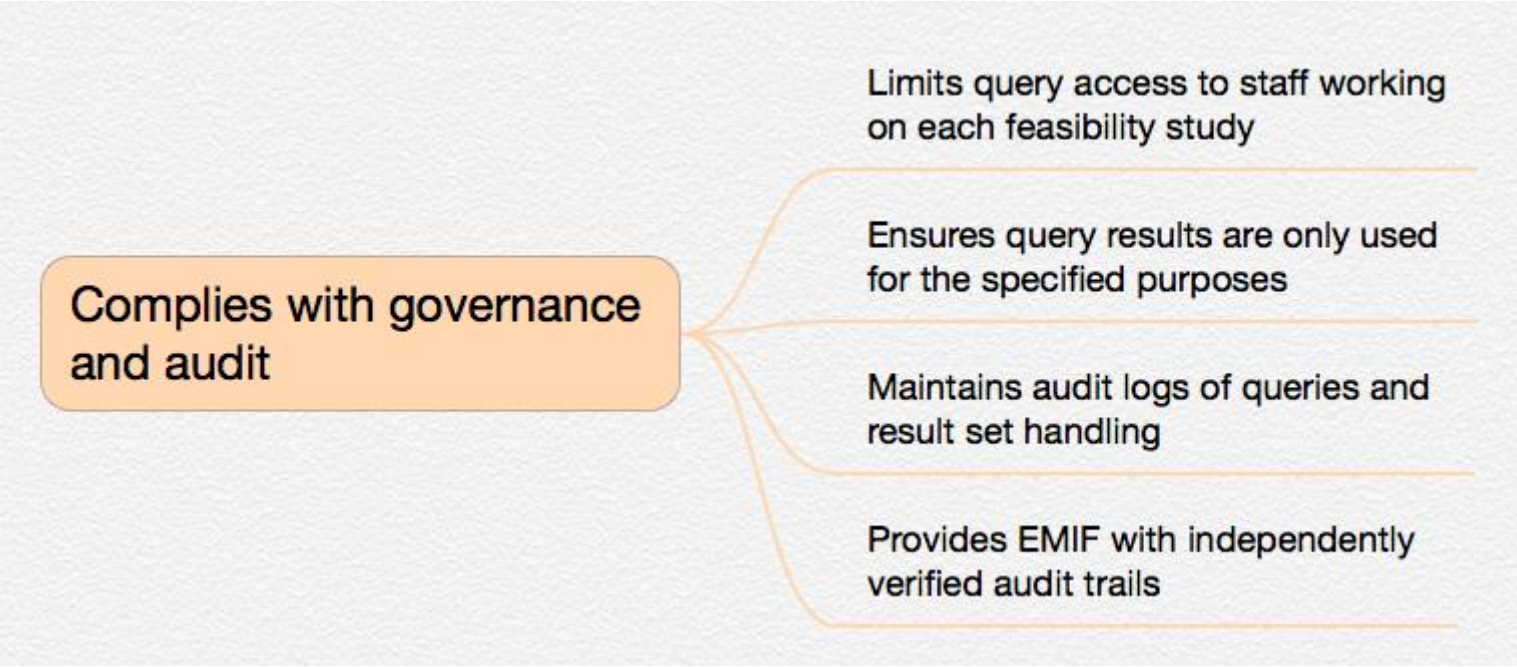


Feasibility: data user obligations

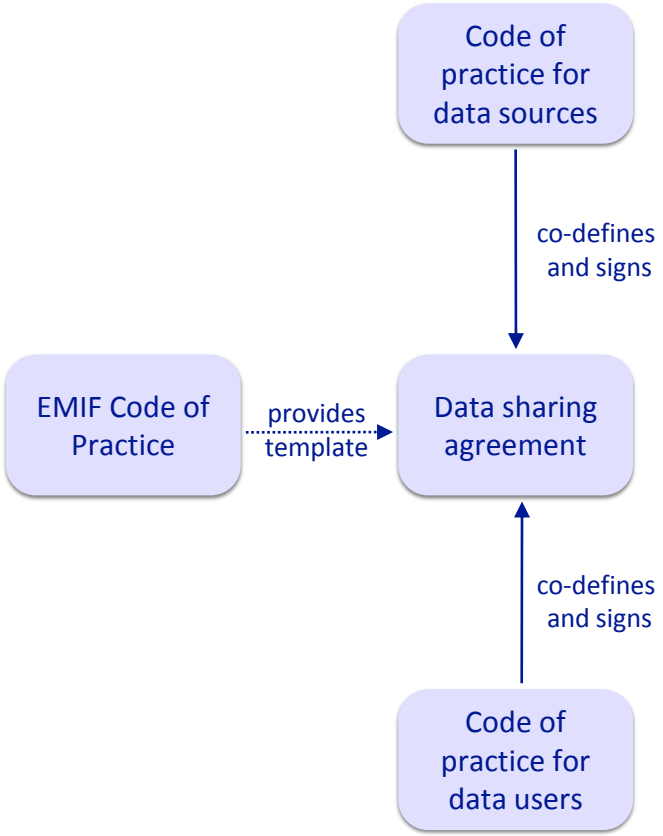


These must match the pre-conditions of each data source on which the query may be executed

Feasibility: data user obligations



Governing EMIF Analysis Services



- Research purposes and protocol
- Data specification and access
- Measures to protect privacy
- Data enrichment services
- Fees to be paid
- Handling of analysis results
- Demonstrating compliance

Putting the ECoP into a bigger context…

- There is a need to:
- champion and govern a trustworthy health data driven ecosystem including EHRs and clinical research platforms
- promote to society the importance of using health data for research, to increase the scale, efficiency and societal benefits from clinical research, to improve health and health care
- engage with society on governance standards that can be jointly upheld by data providers and users, and which are deemed by all to be trustworthy

The European Institute For Innovation Thr~ugh Health Data


EMIF


efpia
imi
innovative
medicines
initiative

LUNCH & DEMONSTRATIONS

Data Harmonisation & Novel Data Reuse

Chair: Assistant Prof Peter Rijnbeek

Erasmus University Medical Center, Rotterdam

OMOP CDM & OHDSI

Assistant Prof Peter Rijnbeek

Erasmus University Medical Center, Rotterdam

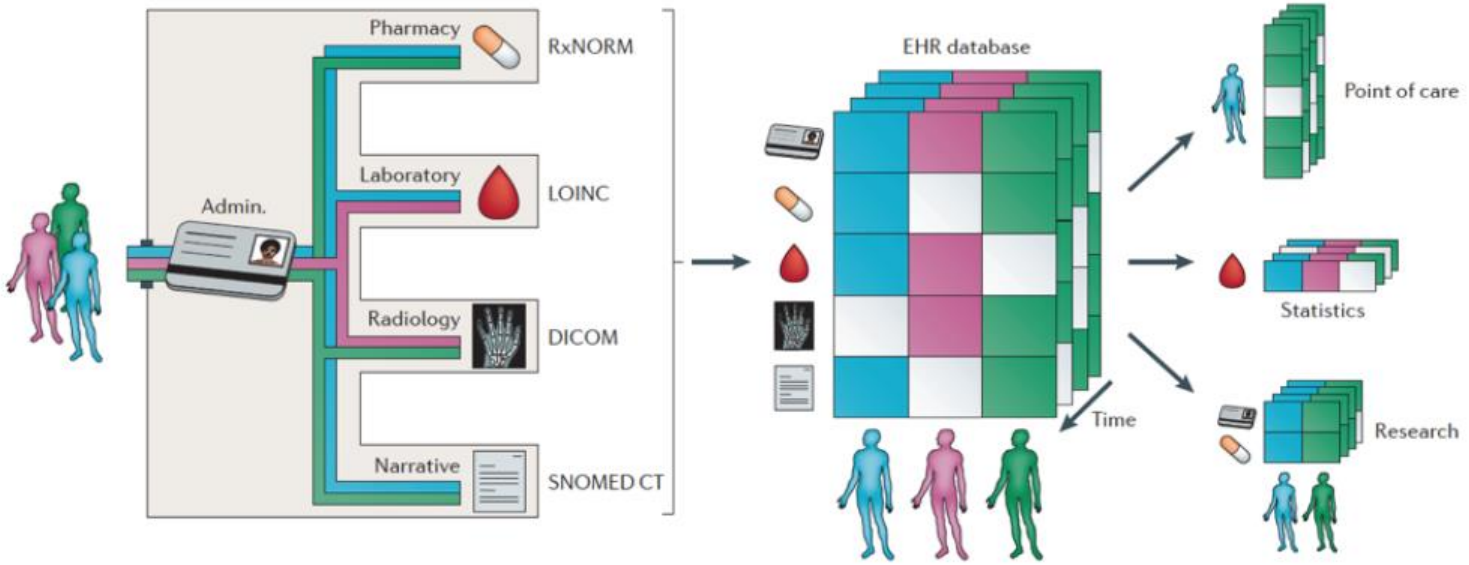
EMIF and the Observational Health Data Sciences and Informatics (OHDSI) initiative

Realising the Value of Health Data ~ Improving Care and Research

September 22th, 2017 Spain

Background

Massive numbers of electronic health records (EHR) are currently being collected globally in observational databases, including structured data in the form of diagnoses, medications, laboratory test results, and unstructured data contained in clinical narratives. This opens unprecedented possibilities for research and ultimately patient care.



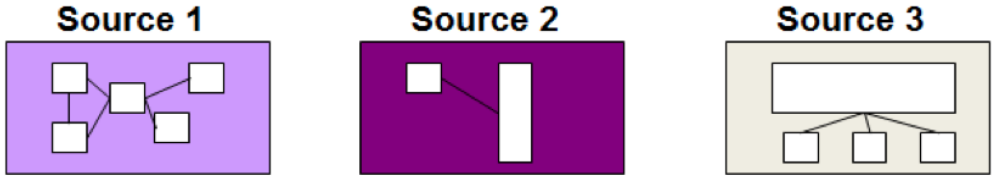
Jensen, Peter B., Lars J. Jensen, and SØren Brunak. "Mining electronic health records: towards better research applications and clinical care." Nature Reviews Genetics (2012).

Challenges

Observational databases differ in both purpose and design. Each has different logical organizations and physical formats, and the terminologies used to describe the medicinal products and clinical conditions vary from source to source.

We need to standardize

Translation to a common data model and standard vocabularies



Any common data model aims to achieve both syntactic and semantic operability.

syntactic operability:
common underlying data structure
(standard grammar)

semantic operability:
common understanding required to interchange information
(standard vocabulary)

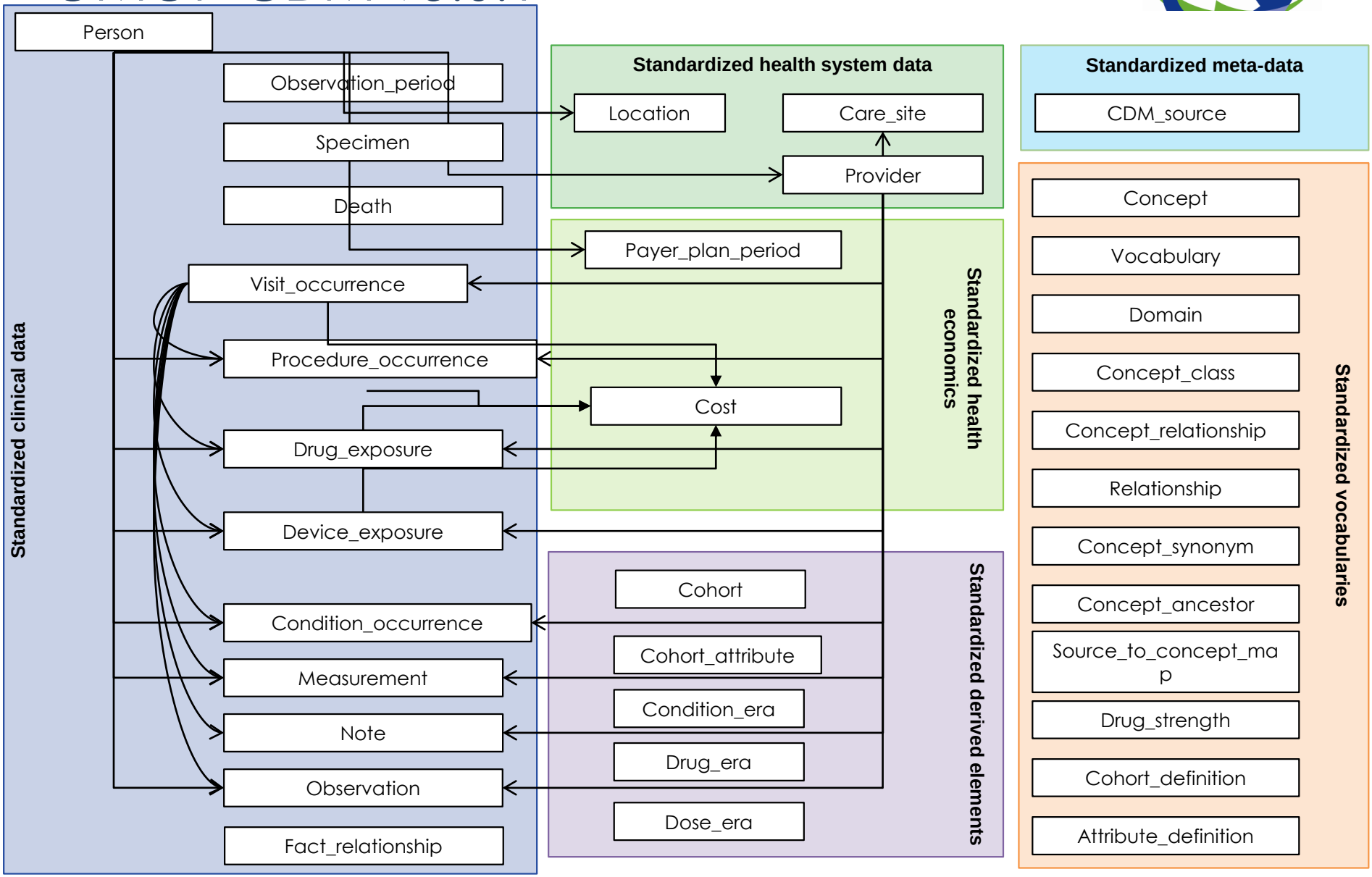
The OMOP CDM and OHDSI

Observational Health Data Sciences and Informatics (OHDSI) has been established as a multi-stakeholder, interdisciplinary collaborative to create open-source solutions for large-scale analytics using the OMOP CDM. http://ohdsi.org

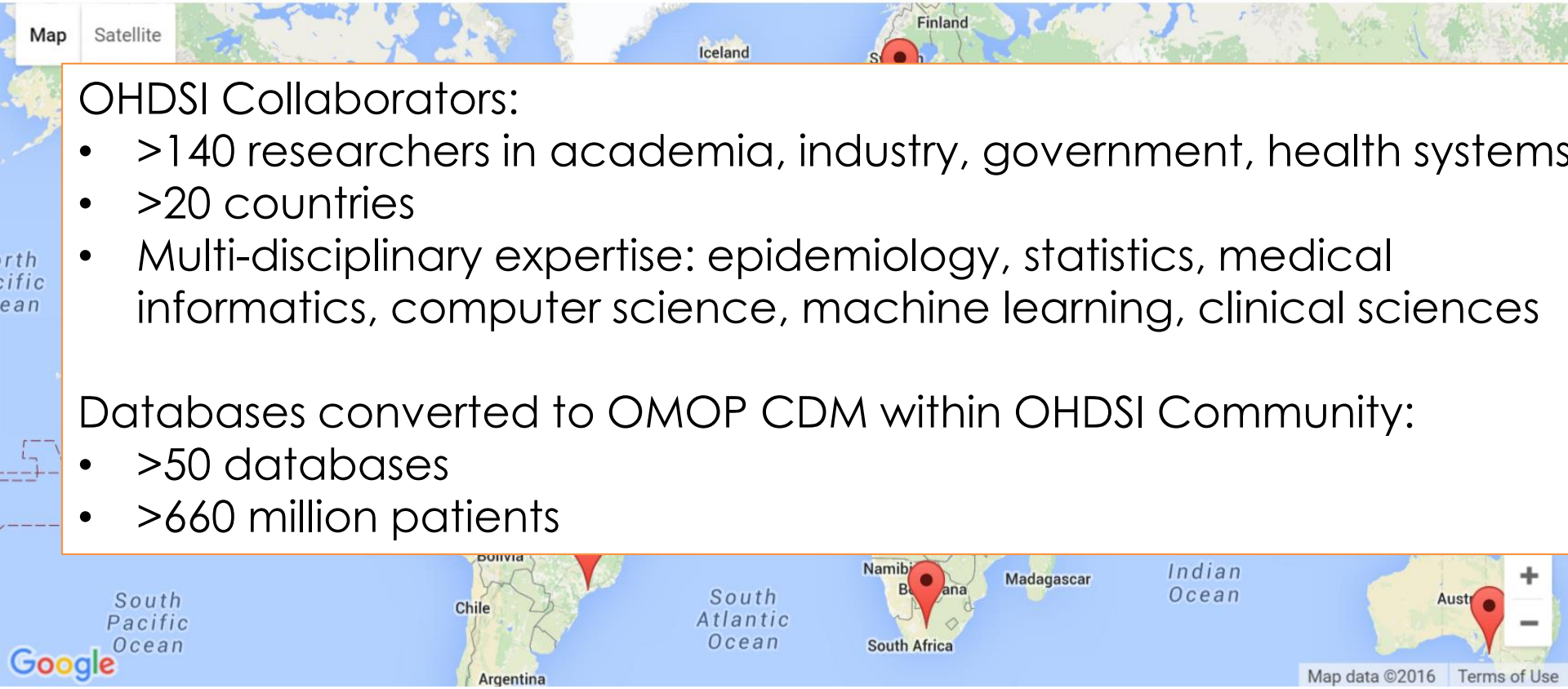
OHDSI has established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University

Hripcsak G, et al. (2015) Observational Health Data Sciences and Informatics (OHDSI): Opportunities for observational researchers. Stud Health Technol Inform 216:574–578.

Deep information model
OMOP CDM v5.0.1



OHDSI community in action



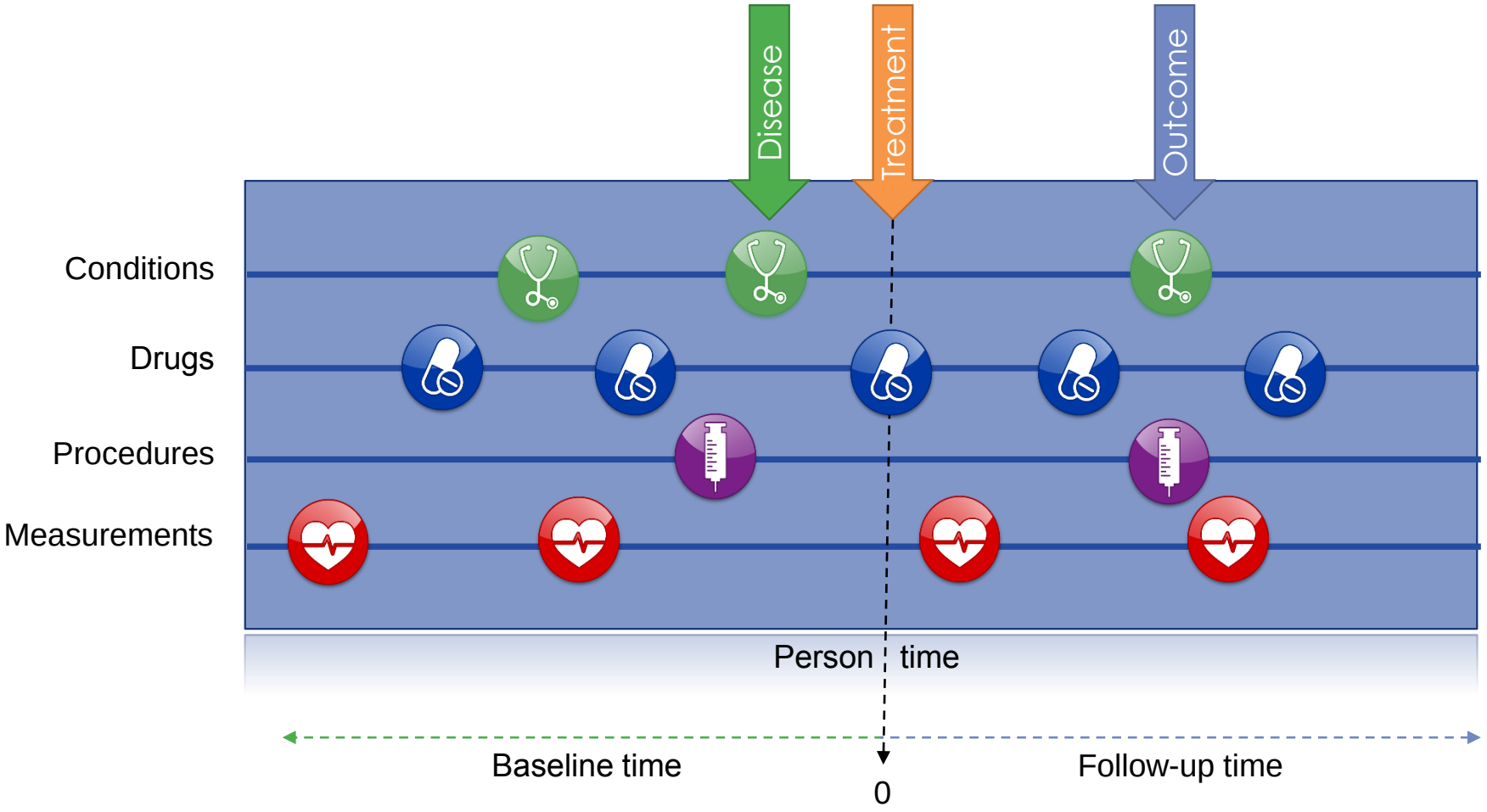
OHDSI Collaborators:

- >140 researchers in academia, industry, government, health systems
- >20 countries
- Multi-disciplinary expertise: epidemiology, statistics, medical informatics, computer science, machine learning, clinical sciences

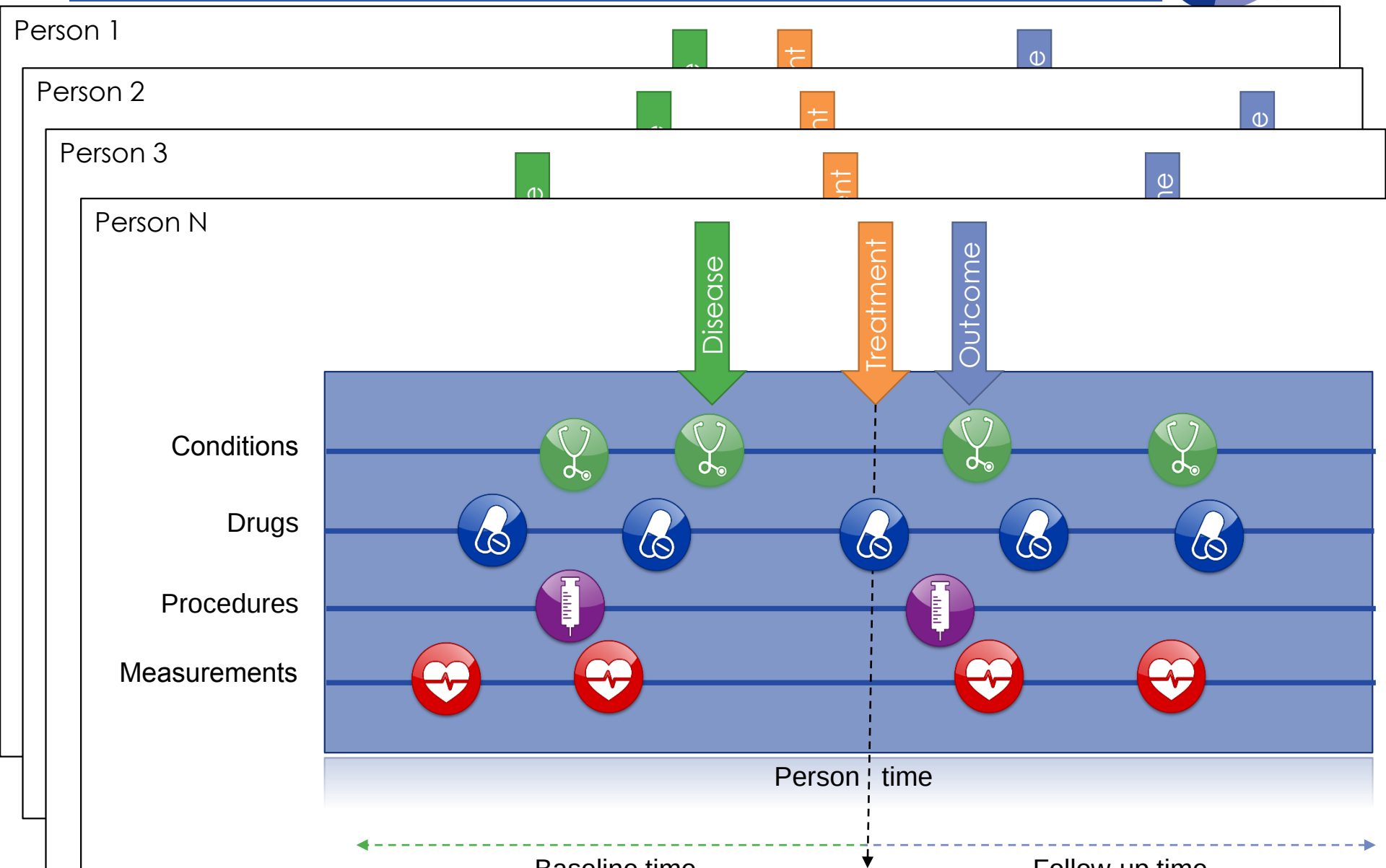
Databases converted to OMOP CDM within OHDSI Community:

- >50 databases
- >660 million patients

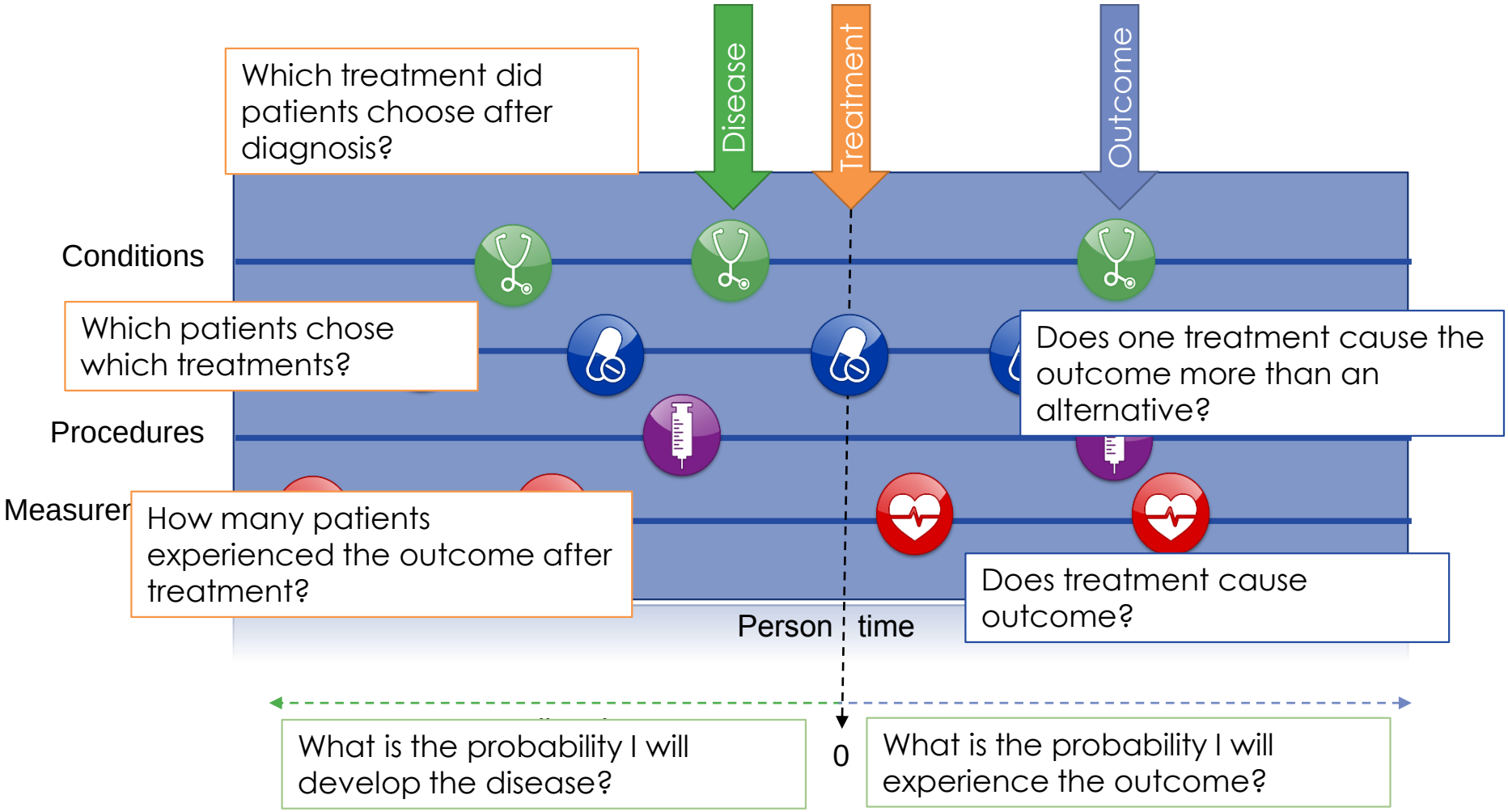
A caricature of the patient journey



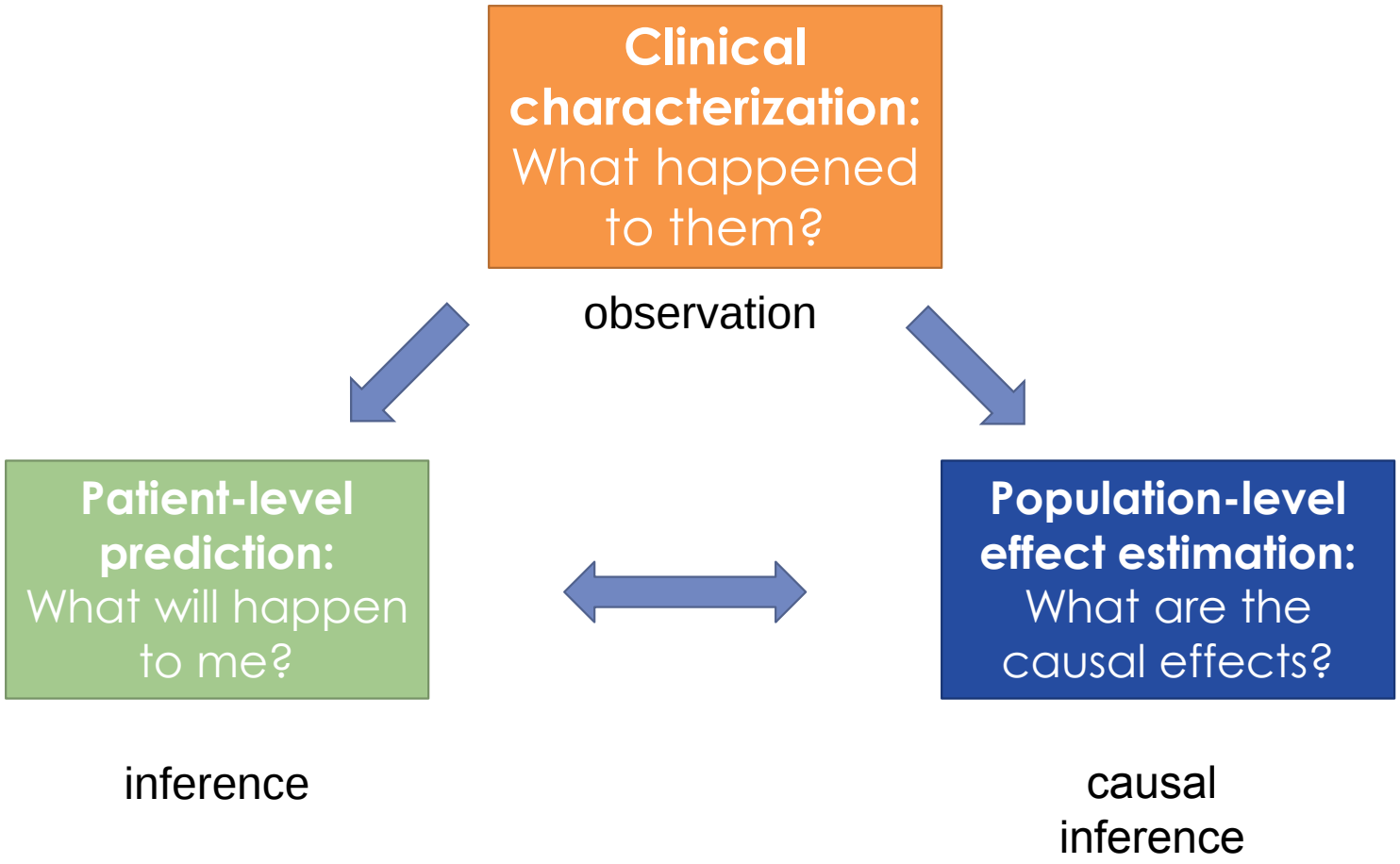
Each observational database is just an (incomplete) compilation of patient journeys



Questions asked across the patient journey



Complementary evidence to inform the patient journey

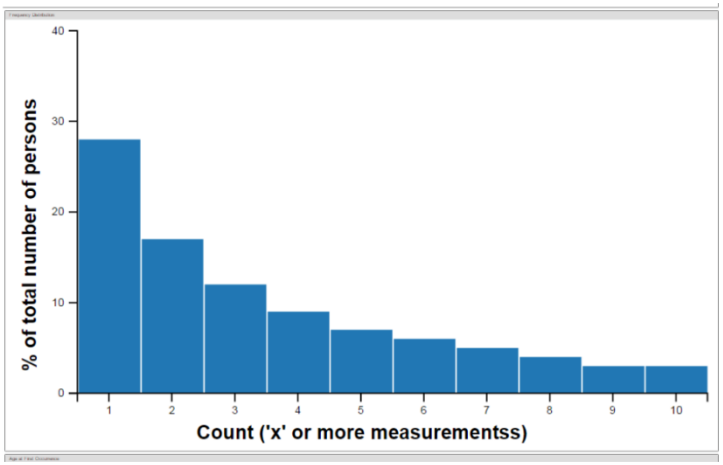


What is OHDSI's strategy to deliver reliable evidence?

- **Methodological research**
  - Develop new approaches to observational data analysis
  - Evaluate the performance of new and existing methods
  - Establish empirically-based scientific best practices
- **Open-source analytics development**
  - Design tools for data transformation and standardization
  - Implement statistical methods for large-scale analytics
  - Build interactive visualization for evidence exploration
- **Clinical evidence generation**
  - Identify clinically-relevant questions that require real-world evidence
  - Execute research studies by applying scientific best practices through open-source tools across the OHDSI international data network
  - Promote open-science strategies for transparent study design and evidence dissemination

Collaboration EMIF and OHDSI

- EMIF has adopted the OMOP-CDM and is actively mapping European databases (see next talk);
- Is incorporating the OHDSI tools in the EMIF Platform;
- Is contributing to the tool development;
- Has supported the addition of security layer on top of the toolset;
- Has evaluated OHDSI tools in the EMIF community



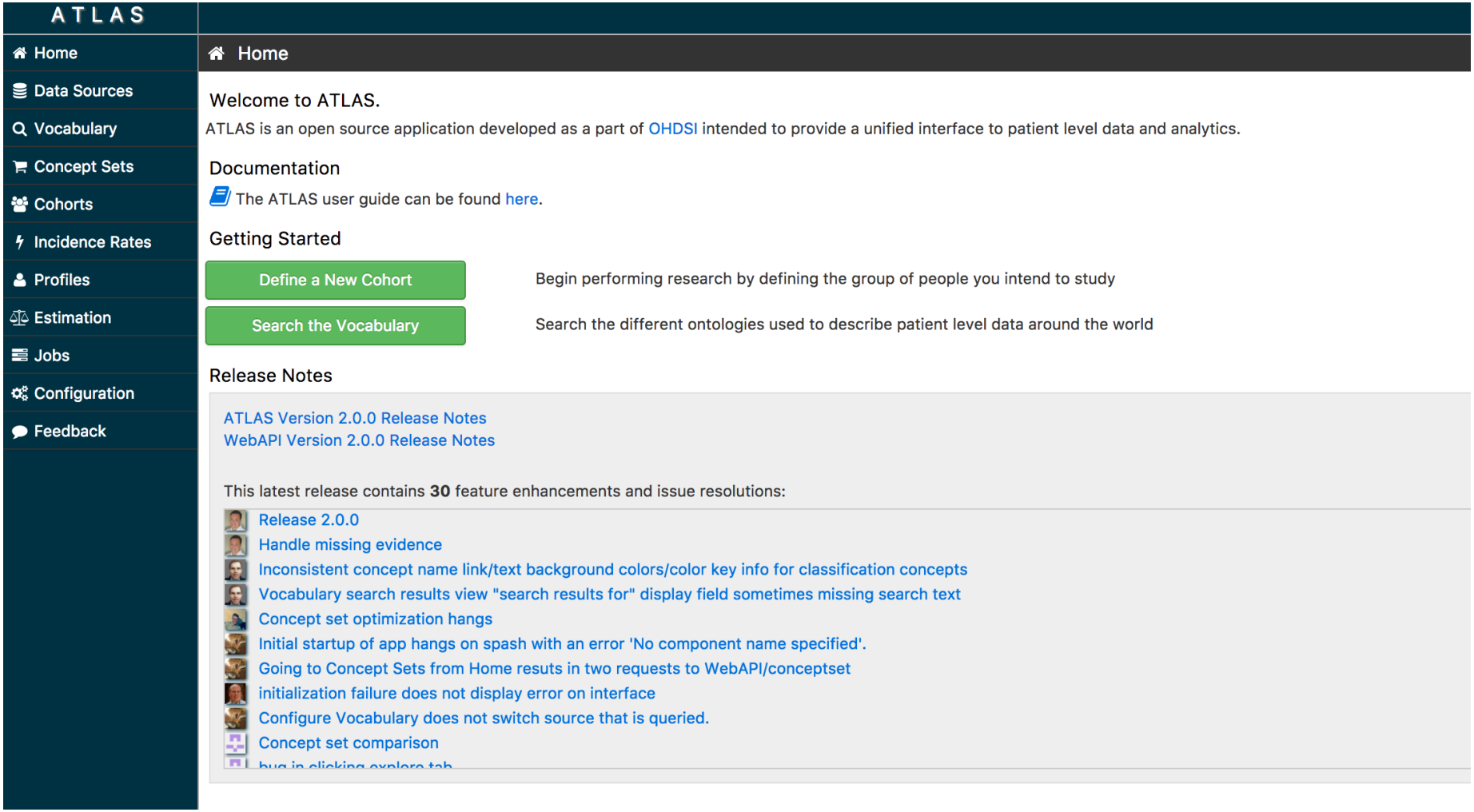
ATLAS

ATLAS is a free, publicly available, web based, open source software tool for researchers to conduct scientific analyses on standardized observational data.

http://www.ohdsi.org/web/atlas (use Chrome)

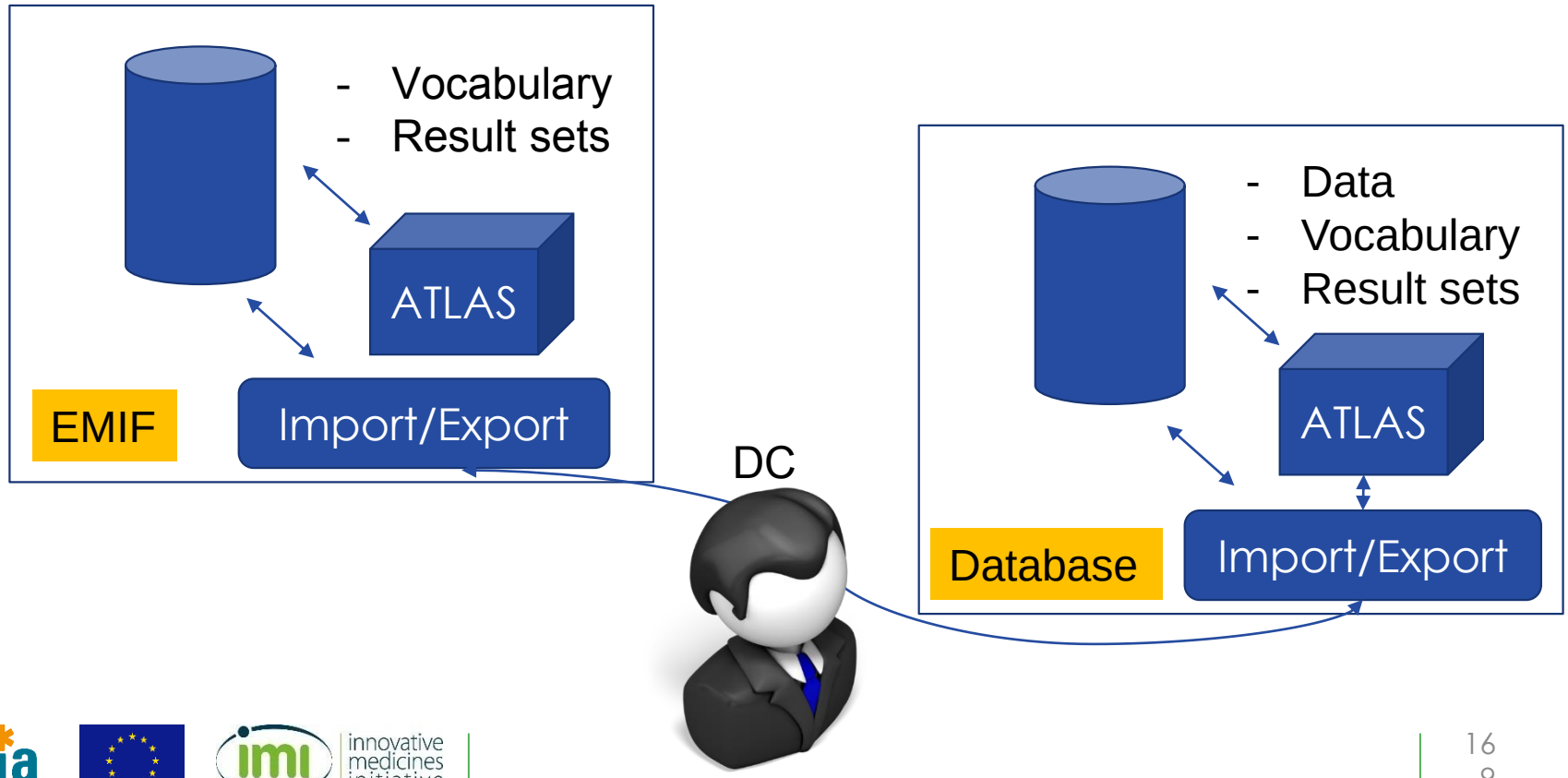
ATLAS

Enabling Research on Standardized data

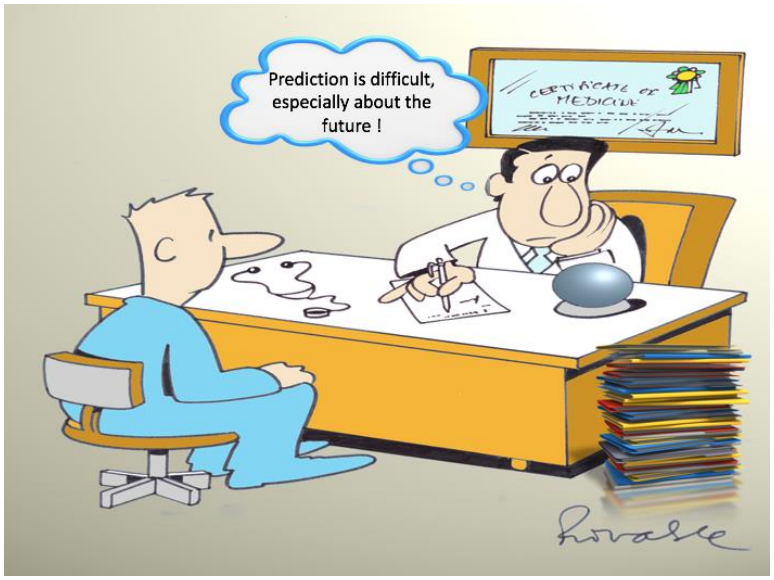


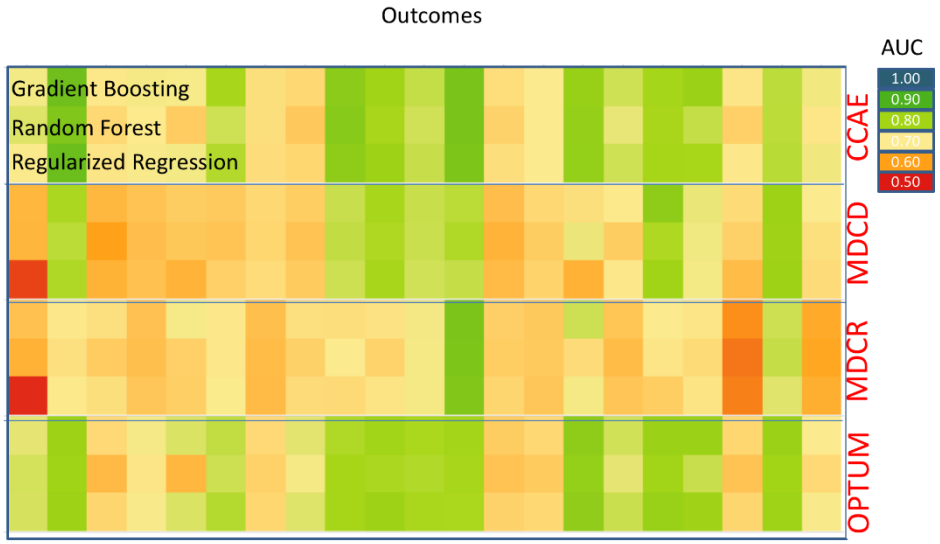
Platform Integration

Integration of Atlas in the EMIF Catalogue -> installation of OHDSI toolset on top of database on central EMIF server

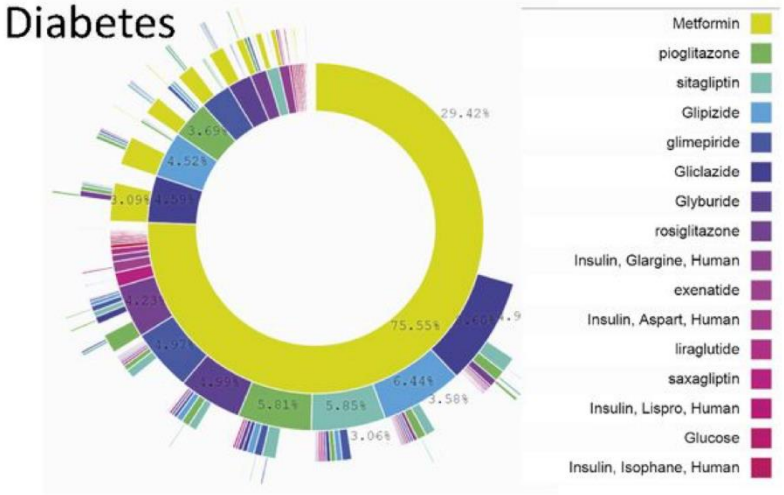


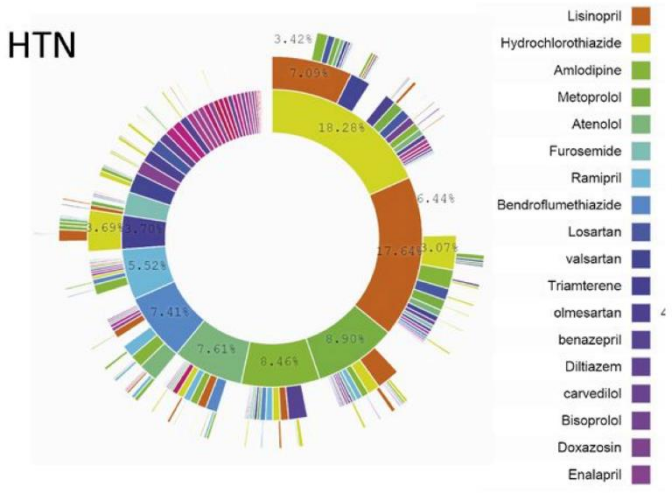
Example:
Large-Scale Patient-Level Prediction





Example: OHDSI Network Study Treatment Pathways





George Hripcsak et al. Characterizing treatment pathways at scale using the OHDSI network **PNAS, 2016 doi:10.1073/pnas.1510502113**

Next Steps

- Continue integration in EMIF platform
- Test runs with feasibility approach
- Treatment Pathways Study in more databases in OHDSI including our EMIF databases with a focus on T2DM
- Workshop with all DCs on the use of the OHDSI tools
- Evaluation of the translation of the European databases to the OMOP-CDM (next talk)


EMIF


efpia
imi
innovative
medicines
initiative

EMIF & Data Custodians Experience with OMP CDM mapping in Europe

Michel Van Speybroeck

Janssen Pharma Data Sciences

EMIF and Data Custodians experience with Observational Medical Outcomes Partnership (OMOP), Common Data Model (CDM) mapping in Europe

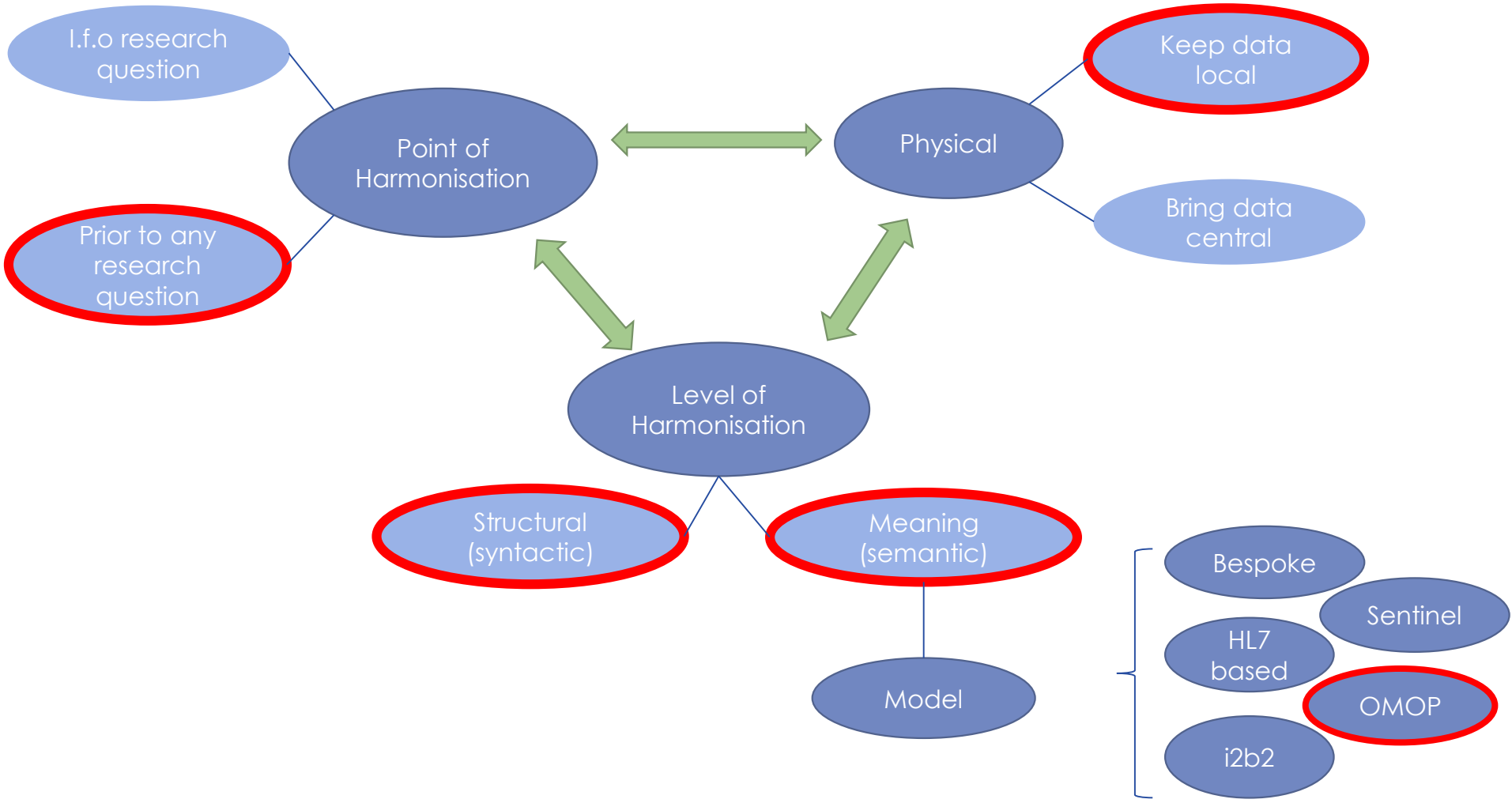
Michel Van Speybroeck - Janssen

September 22nd, 2017

efpia

innovative medicines initiative

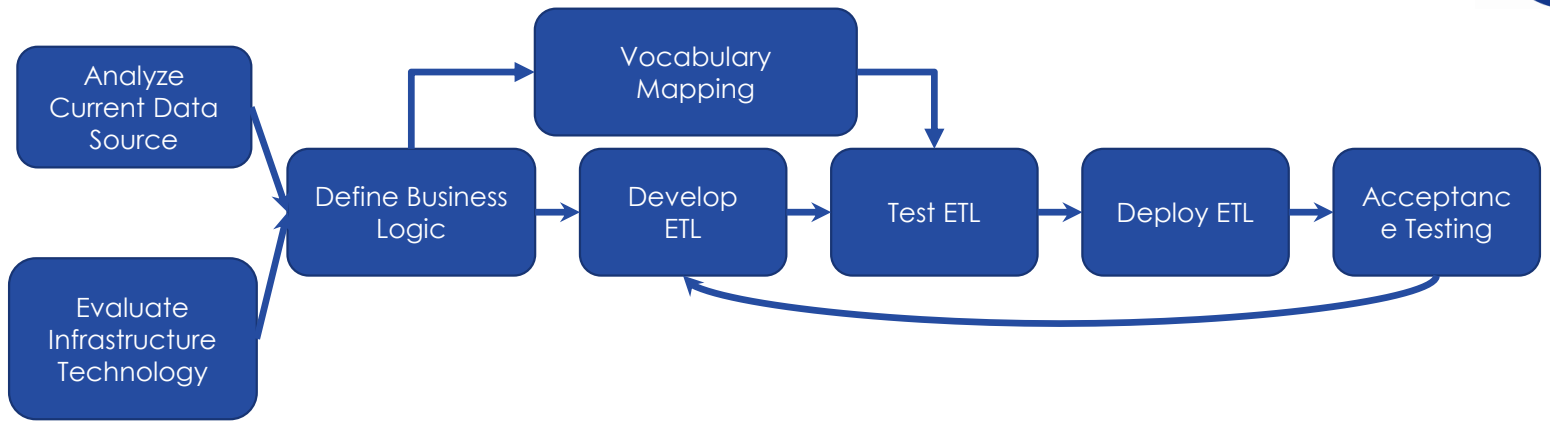
The challenge of health data harmonisation



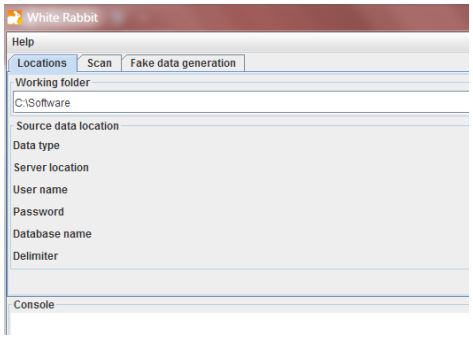
Data sources in scope

Database	Country / Region	Population Size	Type	Mapping Status
Agenzia regionale di sanita della Toscana (ARS	Italy / Tuscany	5. 10^6	Administrative	Completed
Aarhus University Hospital Database	Denmark	2.3 10^6	Administrative	Completed
Health Search IMS Health LPD	Italy	1.6 10^6	Primary care	Completed
Integrated Primary Care Information (IPCI)	Netherlands	2.8 10^6	Primary care	Completed
Pedianet	Italy	0.4 10^6	Pediatric data	In Progress
Pharmo	Netherlands	8.4 10^6	Primary care	Completed for cohort
Information System of Parc de Salut Mar (IMASIS)	Spain	1.4 10^6	Hospital data	In Progress
The Information System for the Development of Research in Primary Care (SIDIAP)	Spain / Catalonia	6.4 10^6	Primary care	In Progress
The Health Informatics Network (THIN)	United Kingdom	12 10^6	Primary care	Completed
Estonian Genome Center at the University of Tartu (EGCUT)	Estonia	52 10^3	Biobank	Completed

The process that was followed



White Rabbit

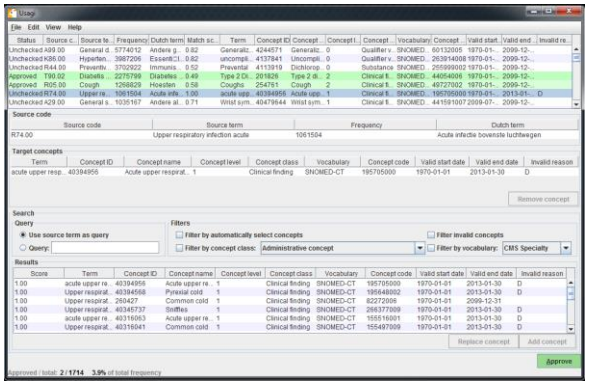


- Profiling of data
- Generating fake data sets

Rabbit in a Hat

- Specification

Usagi



- Vocabulary Mappings

Critical Success Factors

- Bringing the right expertise together:
  - Deep understanding of the source database
  - Understanding of the OMOP CDM structure and vocabularies
  - Technical expertise:
    - Database(s)
    - Extract - Transform - Load (ETL) development – programming language irrelevant
    - Tool installation (OHDSI tools are predominantly based on Java)
- Development of the vocabulary mappings is the most resource intensive activity.
- Focused effort – importance of project management and proper resource allocation
- Quick assessment of results

Effort of mapping to OMOP CDM

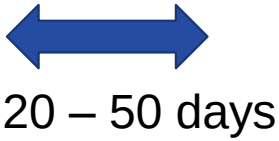
THROUGHPUT TIME:

Total Turnaround time:

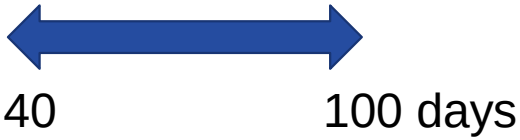
EFFORT:

Preparation (profiling / spec)

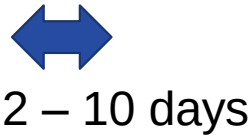
ETL Development



Vocabulary Mapping

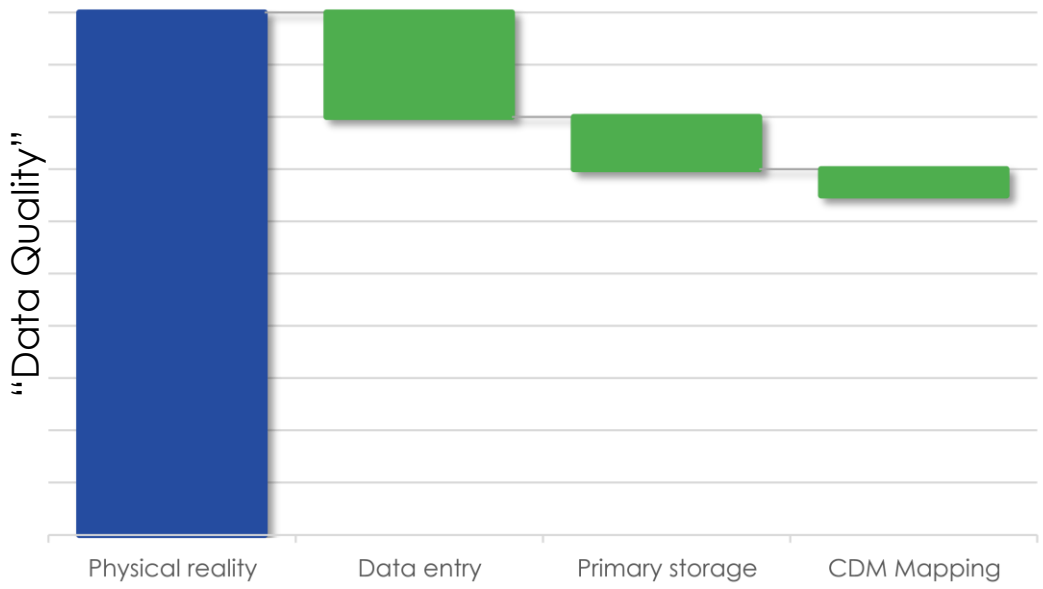


Tool installation

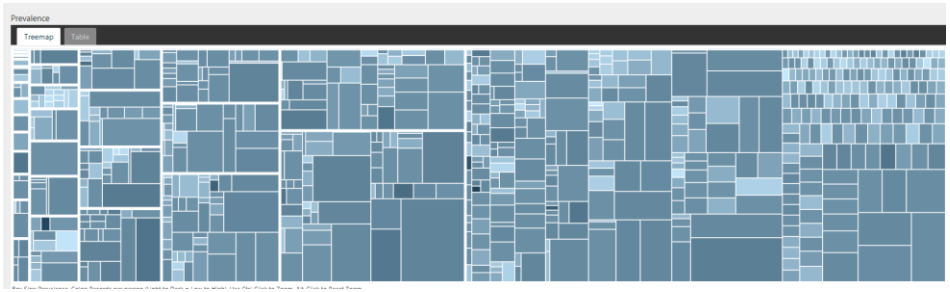


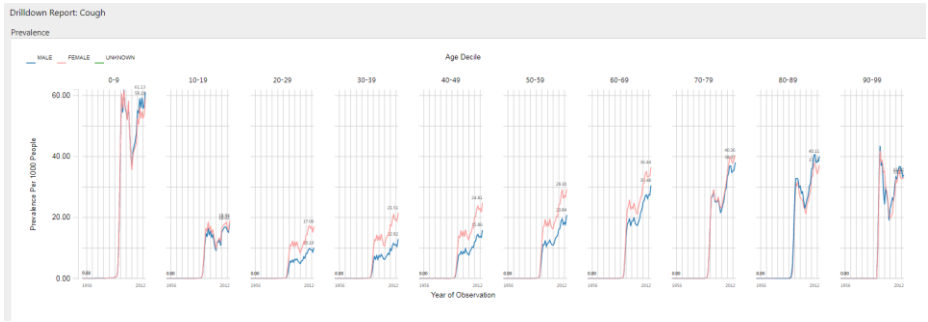
Data Quality and Harmonisation

- Data quality: the degree to which data represent physical reality for a person at a point in time
  - Data accuracy recording
  - Annotation (coding , description, method)
  - Time representation



Verifying Results - Achilles





Verifying Results – Checking Mappings

Example : Drug Level Mappings

Database	Ingredient	Clinical Drug Comp	Clinical Drug Form	Quant Clinical Drug	Clinical Drug	Unmapped
Data Source 1	5%	11%	12%		72%	
Data Source 2	81%					19%
Data Source 3	100%					
Data Source 4	35%	4%	1%		56%	4%
Data Source 5	100%					
Data Source 6	8%			2%	70%	21%
Data Source 7	12%	7%	3%		65%	14%
Data Source 8	100%					

- All relevant source records should be mapped
- Depending on the source between 80 and 100% of codes can be mapped
- Level of mapping might not correspond to the level of source data

Key take-aways

- Participants recognize the benefit of mapping to a common data model
  - Makes the knowledge of the source more explicit
  - Enables scalable research
  - More transparency in protocols
- But even with a CDM you need to have the direct interaction with data custodians to understand elements that are not captured in a data model.
- Mapping to an OMOP CDM is only the first step in a process
- Performing the mappings is a significant effort: dedicated resources, time-boxed and with the right expertise is critical
- A (more) formal process for evaluating the mapping results is required


EMIF


efpia
imi
innovative
medicines
initiative

Working with Cohorts: Switchboxes & Knowledge Objects

Rudi Verbeeck

Janssen Pharma IT

Deep semantic harmonization of clinical cohort data

Rudi Verbeeck

i~HD and EMIF joint event

22 September 2017 – Madrid, Spain

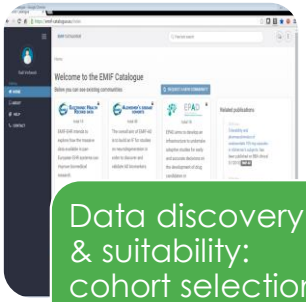
Research cohorts – the supply side

- Deep phenotyping based on research protocol
- Informed consent
- Cohort datasets look similar, but are not the same

Brøndby Haveby, Denmark

Researchers – the demand side

User

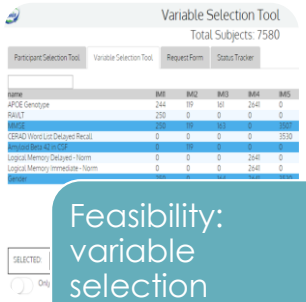


Data discovery & suitability: cohort selection

- Source metadata
- Aggregated, precomputed statistics & profiles

Feasibility: participant selection

- Aggregated results
- Combinations of variables



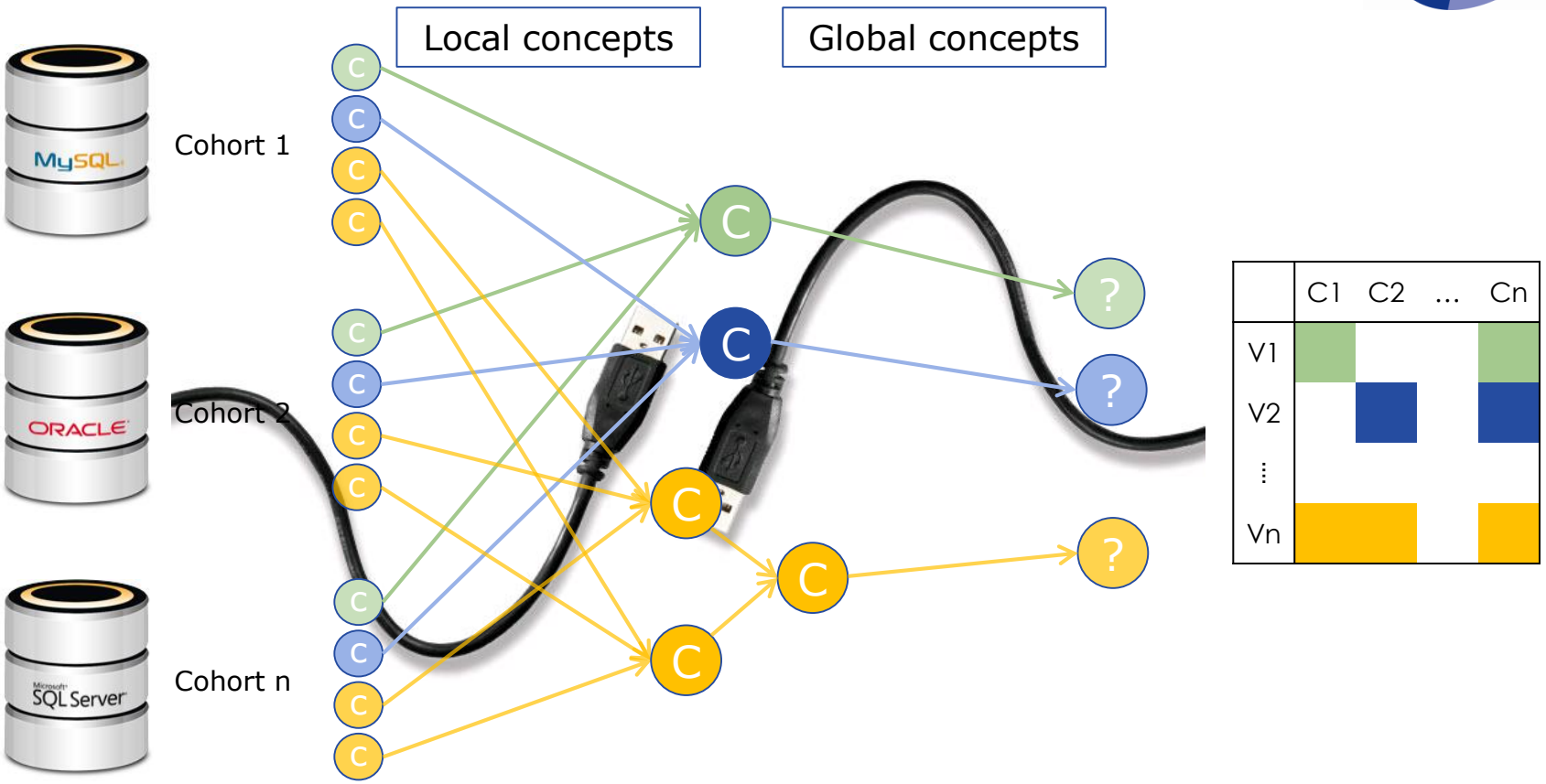
Feasibility: variable selection

- Aggregated results
- Descriptive stats

Analysis: PRRE

- Harmonized subject level data
- Data sharing agreements, ethical approval
- Limited availability

Data harmonization



Data custodians
- Identify local concepts
- Specify mappings
- Define security

Community
- Specify global and derived concepts
- Define research groups

Guiding principles

Generalization

- Compatibility = "inherent quality" + protocol
- Treat mappings, metadata and data equally
- Allow complex mappings

Efficiency

- Distribute ownership knowledge = responsibility
- Build library of reusable objects
- Technical ≠ semantic harmonization

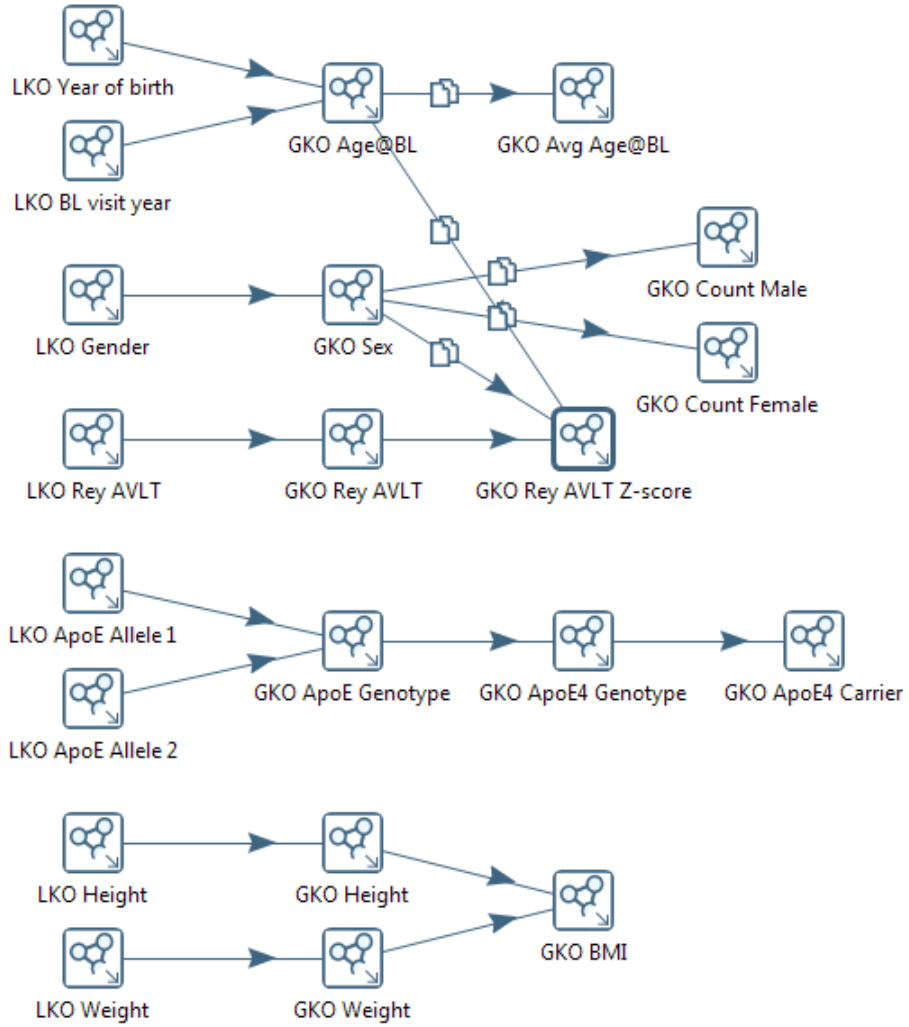
Security

- Set local, propagate to global
- Fine grained
  - Use groupings
  - Use reasoner
- Allow traceback to source data

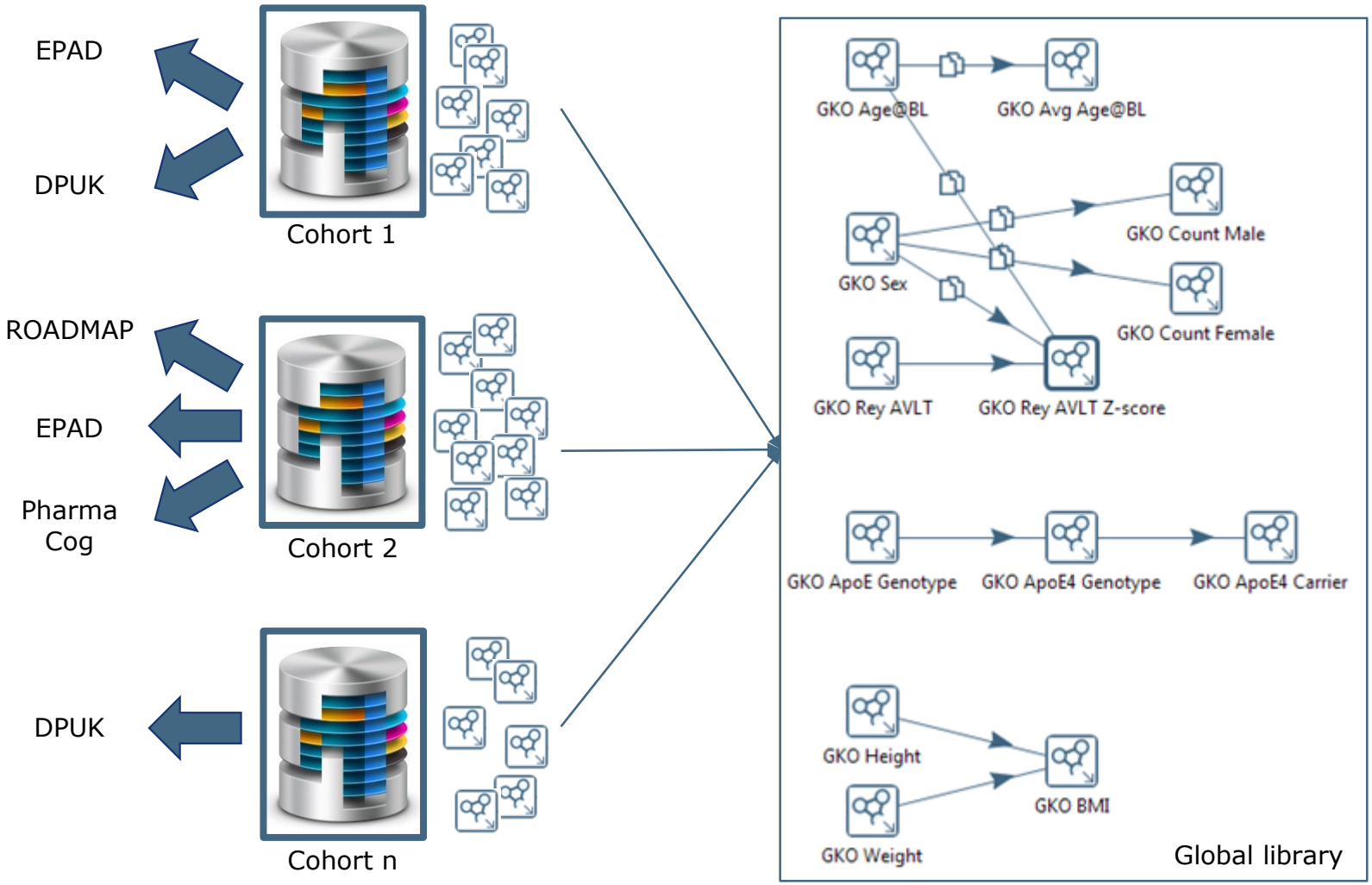
Implementation: semantic web

- Ontology describes application domain
- Specify minimum required information
- Use inferencing (rules) to populate with data

Dependency graph knowledge objects



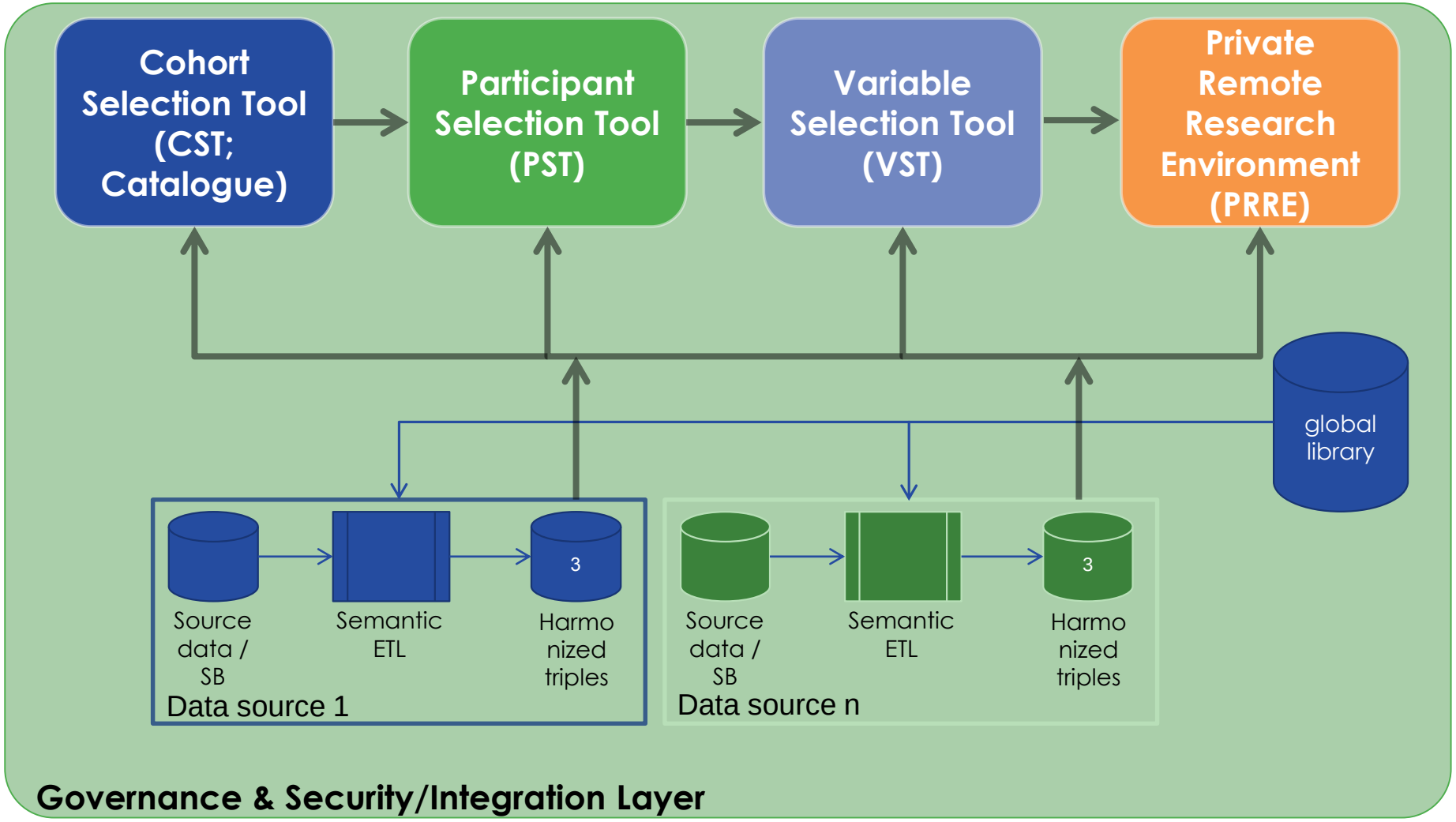
Switchbox


EPAD
DPUK
Cohort 1
ROADMAP
EPAD
Pharma Cog
Cohort 2
DPUK
Cohort n
GKO Age@BL
GKO Avg Age@BL
GKO Sex
GKO Count Male
GKO Count Female
GKO Rey AVLT
GKO Rey AVLT Z-score
GKO ApoE Genotype
GKO ApoE4 Genotype
GKO ApoE4 Carrier
GKO Height
GKO Weight
GKO BMI
Global library

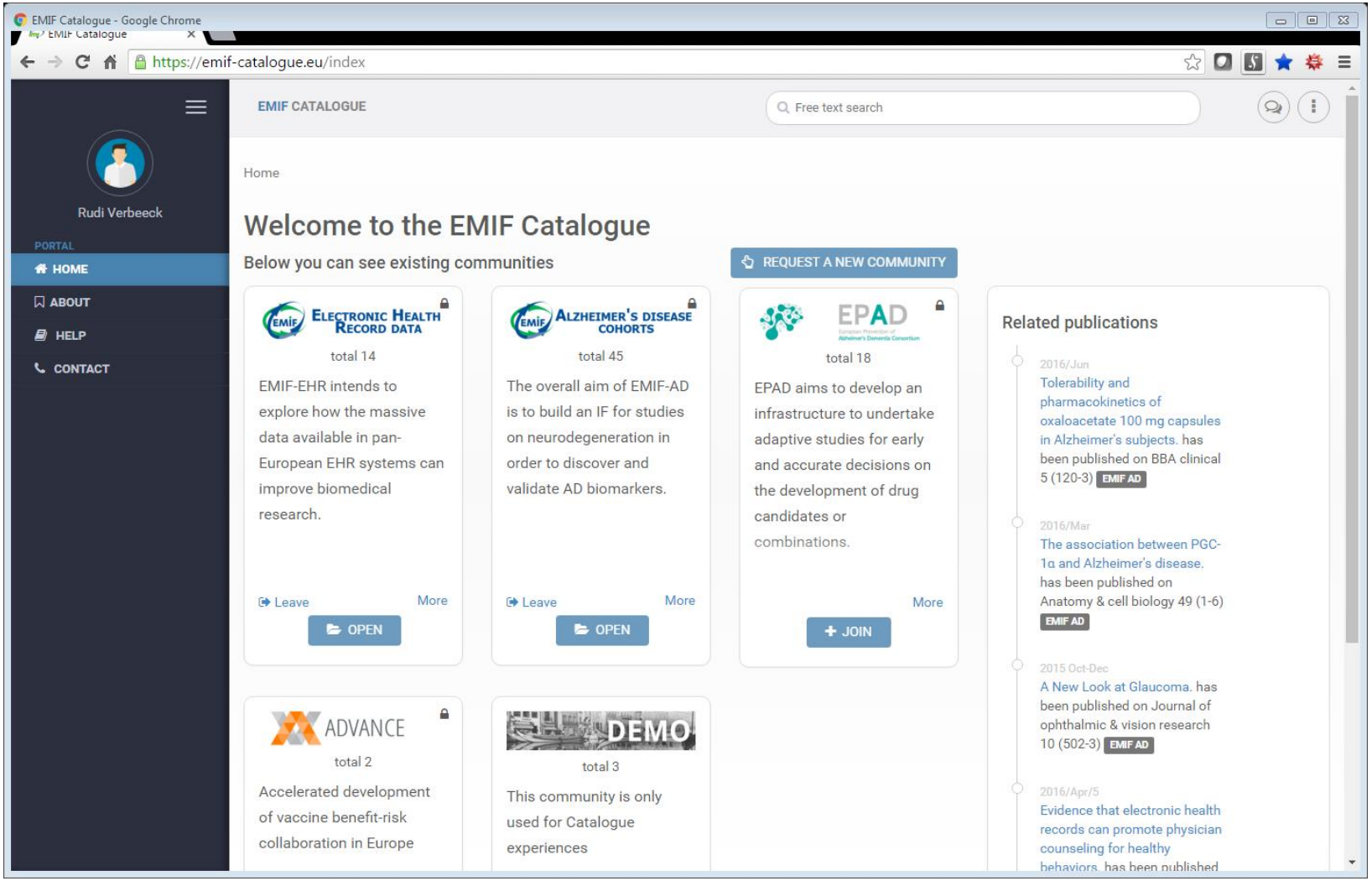
Switchbox

- Present uniform data interface to different projects
- Modelled on the OMOP CDM
- Switchbox contains predefined list (extendible) of harmonized variables
- Automatic connector from Switchbox to knowledge objects (global library)
  - Downstream knowledge objects come for free
  - No mappings from local knowledge objects

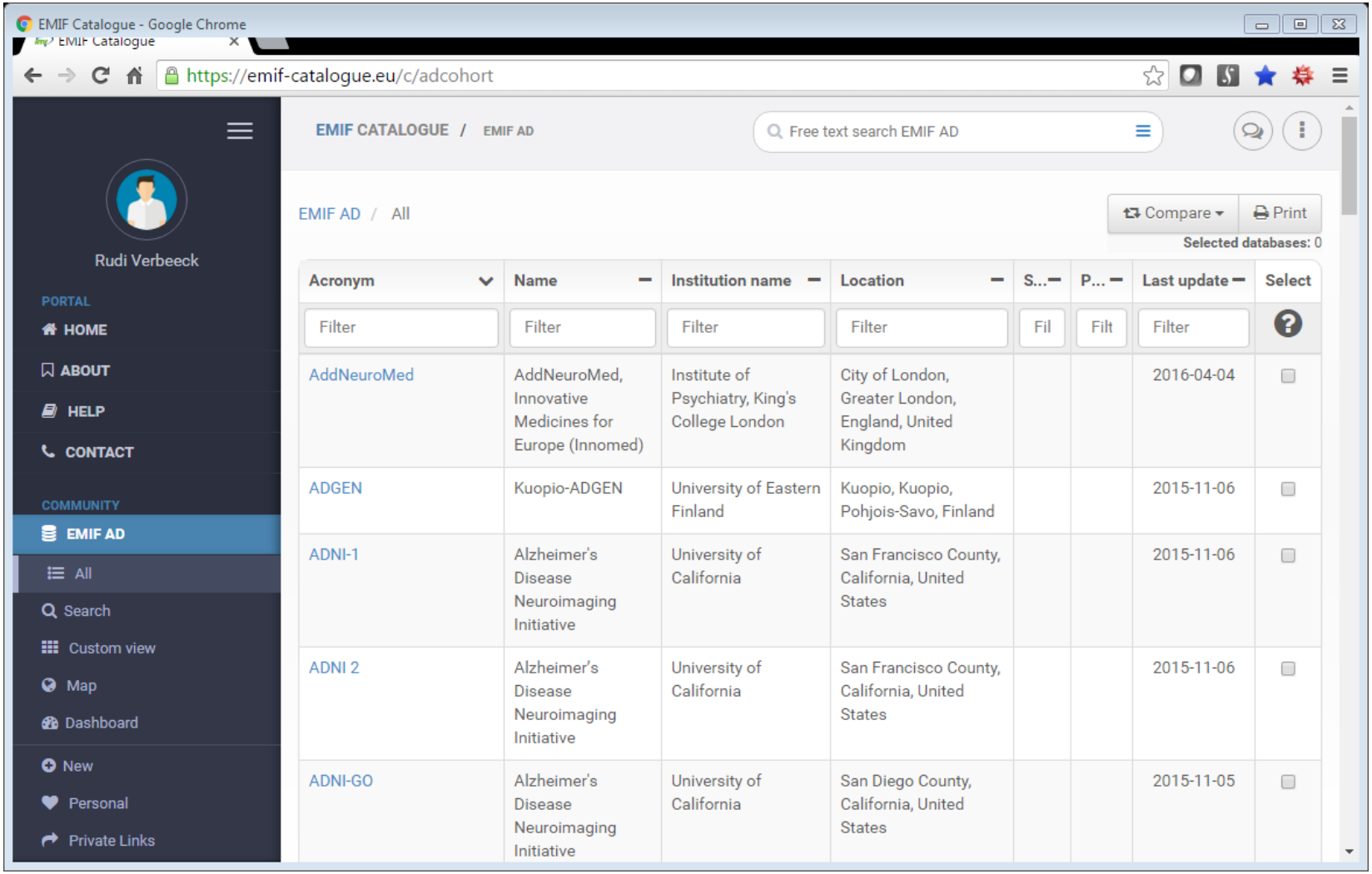
Architecture



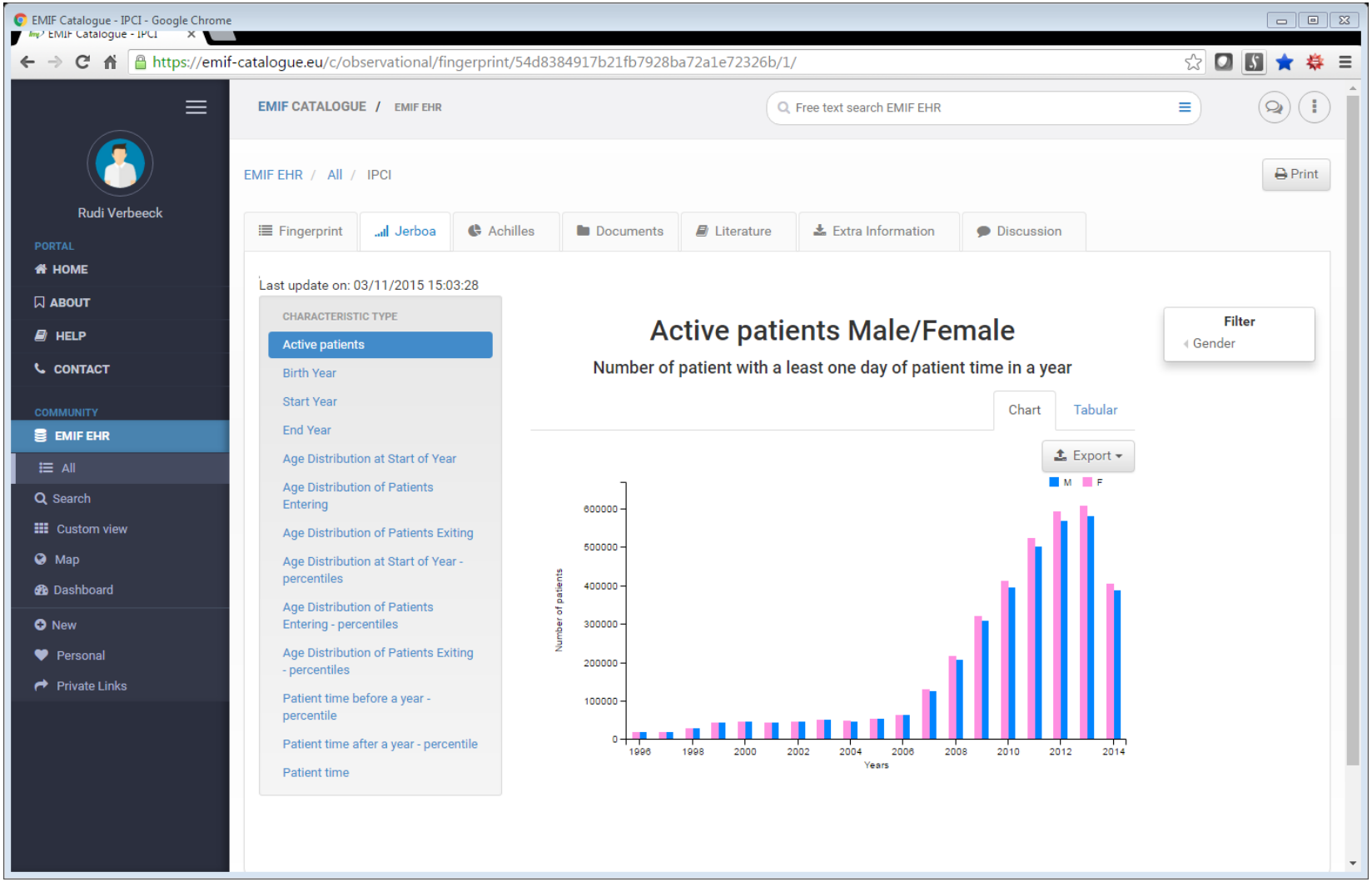
Catalogue - Communities



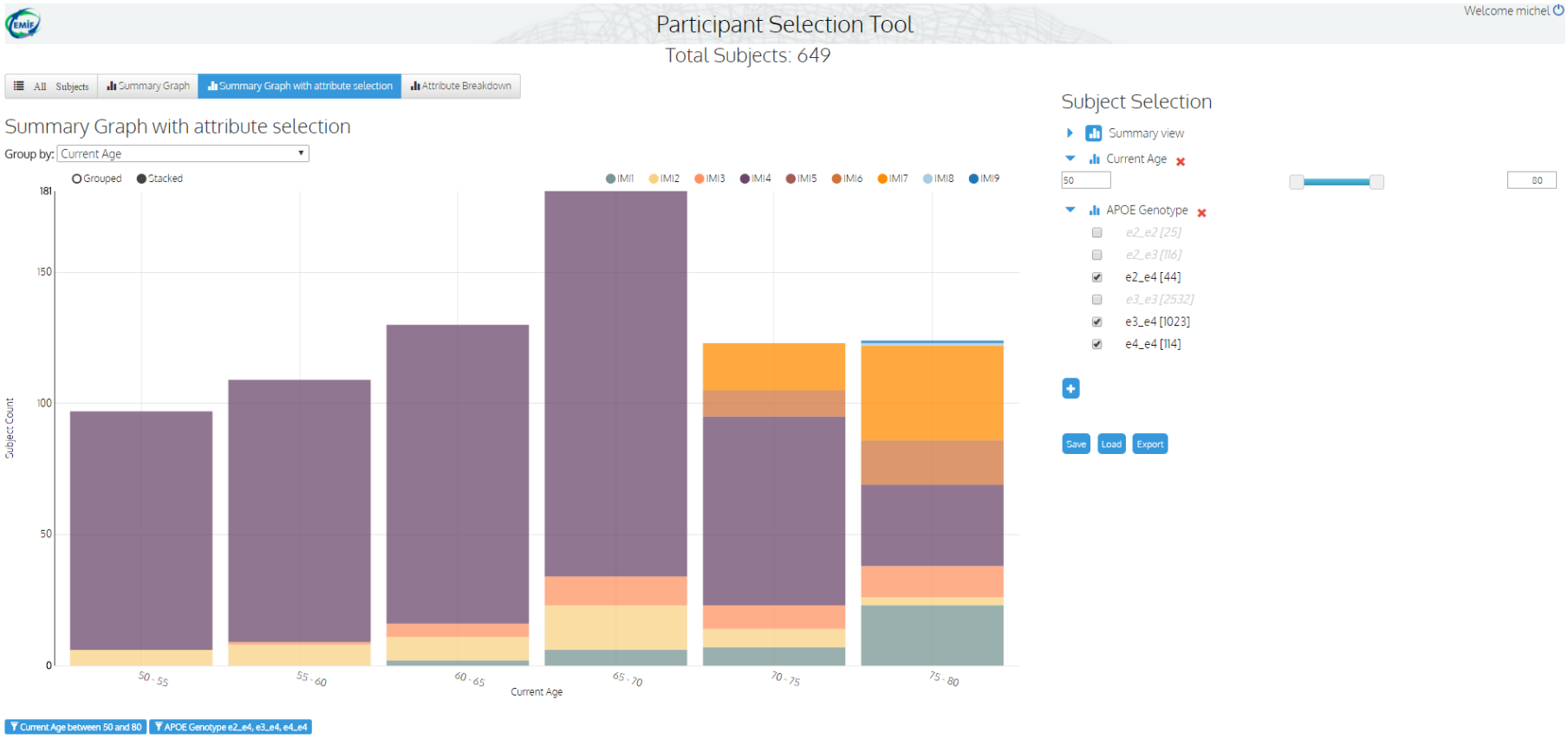
Catalogue – AD cohorts



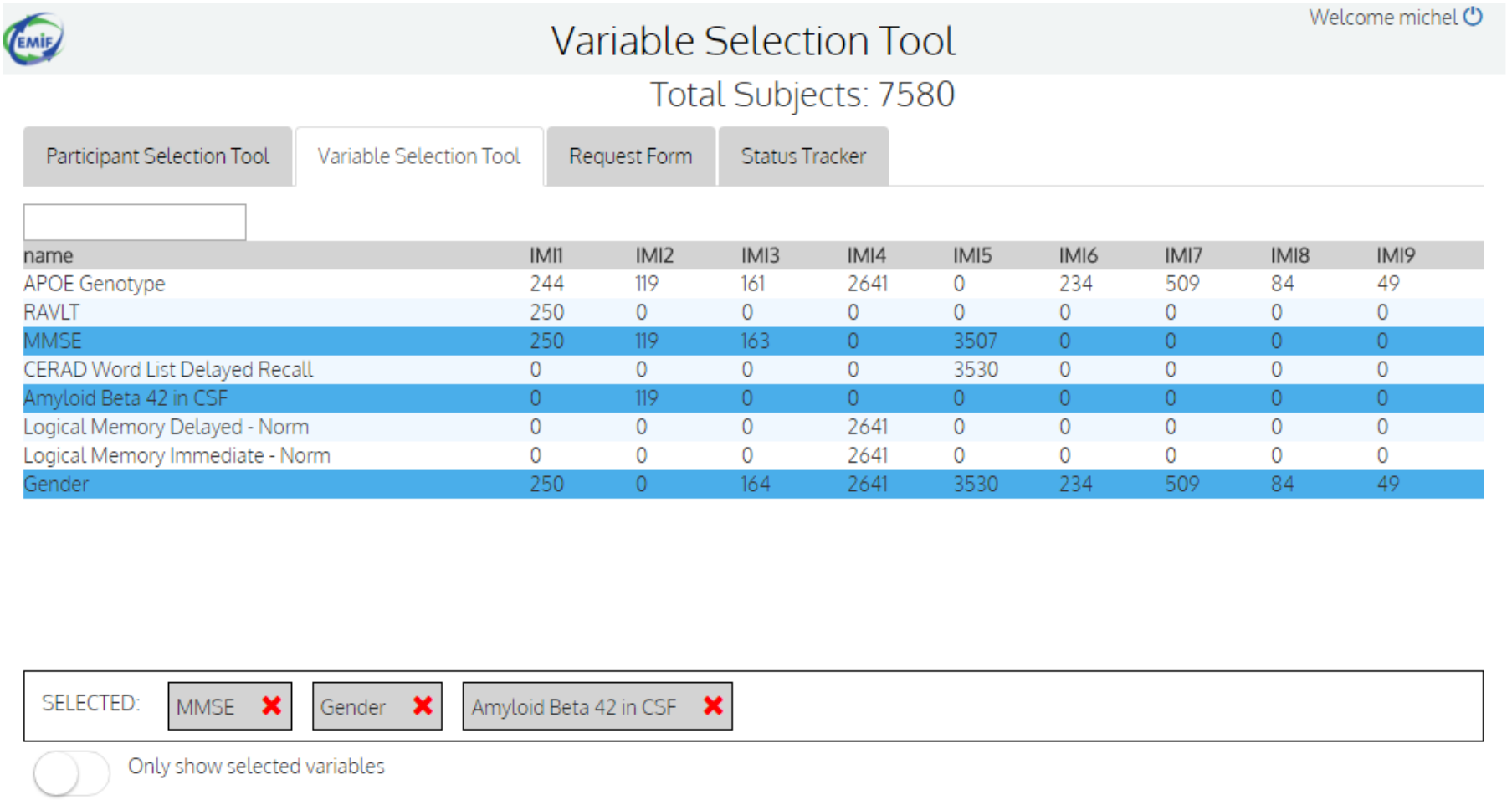
Catalogue – suitability



Participant selection tool

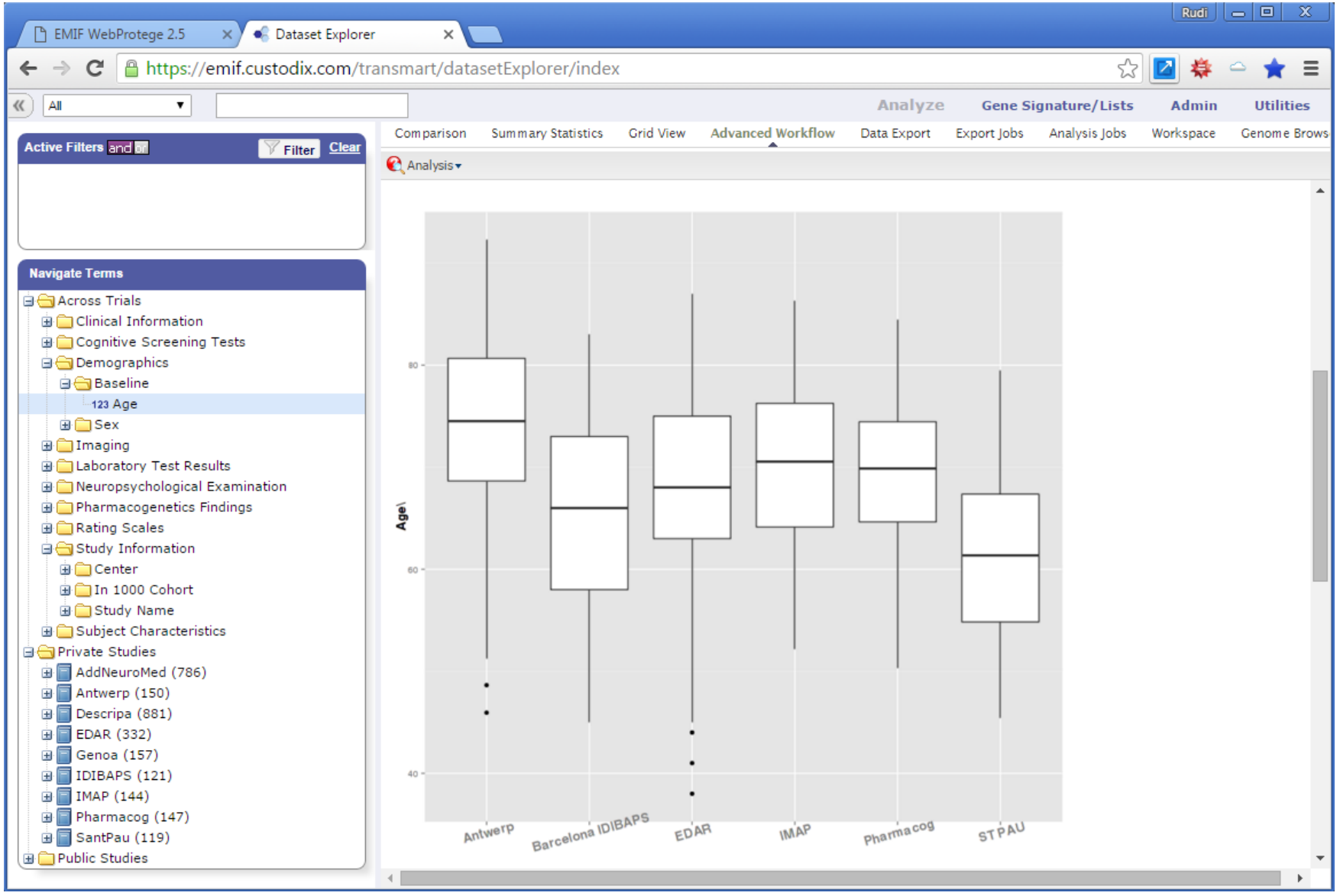


Variable selection tool



name	IMI1	IMI2	IMI3	IMI4	IMI5	IMI6	IMI7	IMI8	IMI9
APOE Genotype	244	119	161	2641	0	234	509	84	49
RAVLT	250	0	0	0	0	0	0	0	0
MMSE	250	119	163	0	3507	0	0	0	0
CERAD Word List Delayed Recall	0	0	0	0	3530	0	0	0	0
Amyloid Beta 42 in CSF	0	119	0	0	0	0	0	0	0
Logical Memory Delayed - Norm	0	0	0	2641	0	0	0	0	0
Logical Memory Immediate - Norm	0	0	0	2641	0	0	0	0	0
Gender	250	0	164	2641	3530	234	509	84	49

tranSMART cross-trial analysis



Data in tranSMART

Cohort name	# Subjects	# in cross trial	# in 1000 samples cohort	# expected in 1000 samples	# unique variables	# visits
AddNeuroMed	774				46	14
Amsterdam	172	172	172	170	67	14
Antwerp	150	150	150	150	147	14
Città – GAP	40	40	40	40	120	1
Descripa	881	881	29	228	215	12
EDAR	332	332	203		173	6
Genoa	157	157			35	2
Gothenburg	95	95	95	101	127	7
IDIBAPS	164	164	120	120	92	11
IMAP	144	144			85	2
Lausanne	40	40	40	40	112	8
Leuven	180	180	180	180	53	1
Pharmacog	147	147	147	147	59	5
Sant Pau	135	135	45	45	150	8
Total	**3411**	**2637**	**1221**	**1221**		

Conclusions

EMIF key distinguishing features:

Data custodians - Supply

- Empower data owners and SMEs to distribute the workload for deep harmonization
  - Metadata
  - Mappings
- Specify minimum info for unambiguous interpretation of metadata & mappings, generate the data
- Build library to encourage re-use
- Control access

Researchers - Demand

- Progressive protocol refinement and drill down to the data
  - Data discovery
  - Suitability
  - Feasibility
  - Data analysis
- Tools
  - Cohort selection tool
  - Patient selection tool
  - Variable selection tool
  - PRRE

Acknowledgements

Data harmonization framework

- Janssen: Luiza Gabriel, Alvaro Cortes-Callabuig, Michel van Speybroeck
- U. Manchester: James Cunningham

Tools and scripts

- Know.Bi: Bart Maertens

Library

- ITTM: Serge Eifes, Adriano Barbosa, Kavita Rege

TranSMART data load

- Maastricht U.: Isabelle Bos, Stephanie Vos
- The Hyve: Janneke Schoots – van der Ploeg, Olaf Meuwese, Andy Sewgobind, Stefan Payralbe


EMIF


efpia
imi
innovative
medicines
initiative

Placebo as a Surrogate for RWD

Prof Derek Nunez, Gurparkash Singh & Peter Egger

Duke University, US,

Janssen Pharma R&D & RWE & Epidemiology GSK

Re-Use of Clinical Trial Data

A Case Study Sponsored by EMIF-Metabolic

Derek Nunez MD FRCP (presented by Peter Egger PhD, GSK)
& Gurparkash Singh PhD, Janssen

Madrid September 2017

Disease insights from a variety of data sources

- Observational patient health data sources
  - Administrative databases for health insurance purposes
  - EHR data for patient management purposes
  - Disease / treatment registries
  - Biobanks

- Patient health data from Clinical Trials
  - Clinical Trials are conducted to evaluate the safety and efficacy of a new treatment
  - Can Clinical Trial data be re-used to evaluate disease?

Clinical Trial Data from Placebo Arms

Pros

- Trials can be very large (10,000 +) and long (3+ years)
- Placebo arms
  - No Investigational Drug(s) to complicate interpretation
  - Subjects often on ‘Standard of Care’ medications
- Subjects are observed periodically using standardized reporting tools (physical exams, laboratory measurements, ECGs etc)
- Longitudinal trends may be discernable
- May include novel data collections – digital data directly from patients, such as from wearables, real-time recording by patients

Clinical Trial Data from Placebo Arms

Challenges

- Providing ‘real world’ insights
  - Inclusion/Exclusion criteria may skew subjects away from “Real World” patients
  - More intense disease monitoring and management
  - Close observation of subjects may alter behaviour
  - Subjects may drop-out during a trial
- Access to data
  - Trial consent forms must allow the re-use of data
  - May be difficult to collaborate across companies

Case study

Background

- Nonalcoholic Fatty Liver Disease (NAFLD) is commonly associated with obesity and/or type 2 diabetes
- NAFLD is common (10-30% of adults), but progression to more severe liver disease is uncommon and predictors are not well understood

Key objectives

1. How well do BMI and liver endpoints track together?
2. NAFLD progression and baseline predictors

Can Clinical Trial data be re-used to address these objectives?

- Can use data from trials not designed to investigate specifically NAFLD objectives but where NAFLD measures such as Alanine aminotransferase (ALT) and aspartate aminotransferase (AST) are measured as ‘safety biomarkers’

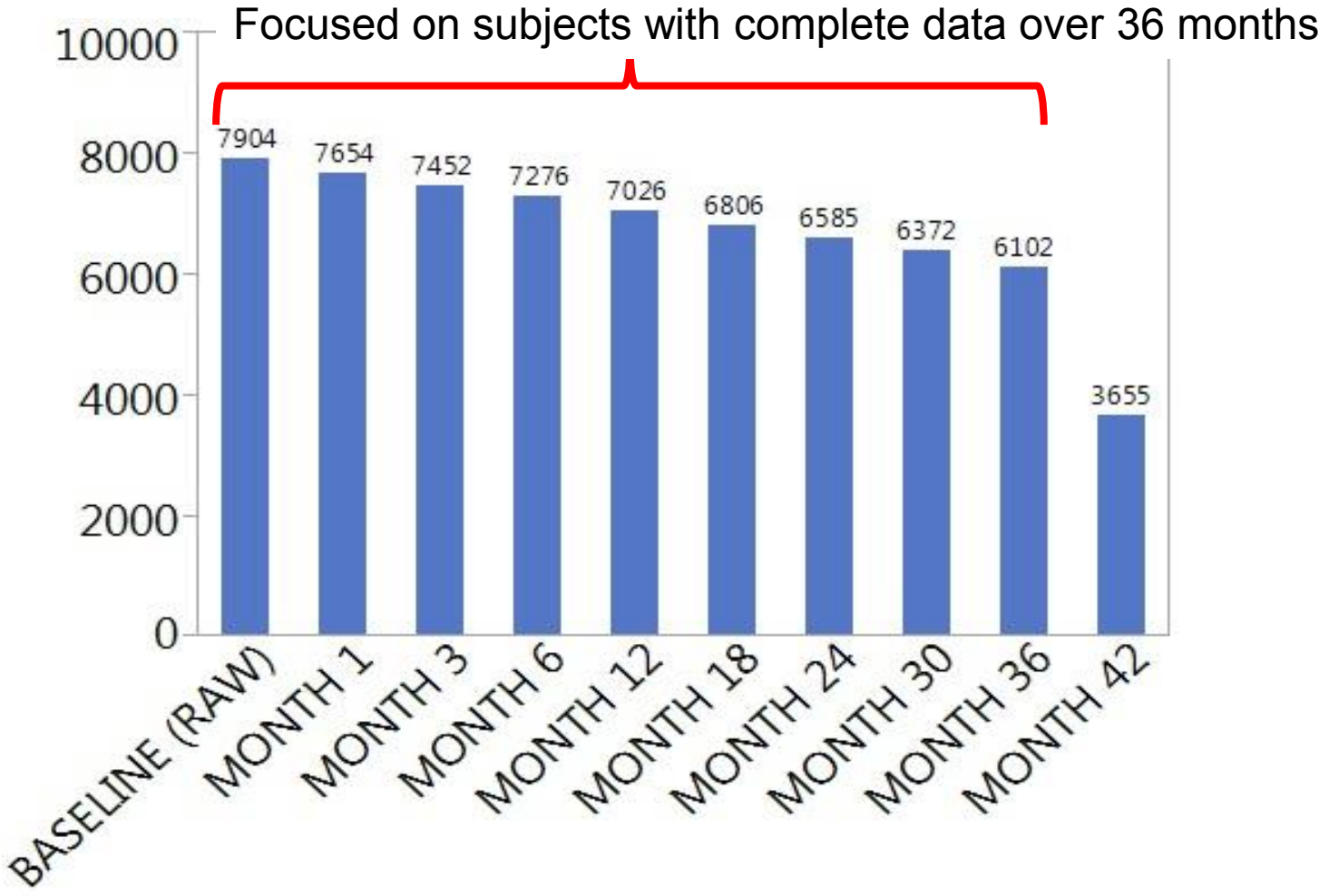
The STABILITY trial (The STabilisation of Atherosclerotic plaque By Initiation of darapLadIb TherapY)

(N Engl J Med 2014;370:1702-1711)

- Tested darapladib (LpPLA$_2$ inhibitor) vs Placebo
- 15,828 subjects enrolled (663 centers in 39 countries)
  - At high cardiovascular risk (chronic coronary artery disease or risk factors [one of age ≥60 years, diabetes, smoker, low HDL-C, polyvascular arterial disease, renal dysfunction])
- High background use of standard-of-care treatment (eg. statin therapy)
- Randomization to darapladib or placebo
- Median duration of follow up: 3.7 years

Cardiovascular risk factors		
Diabetes requiring pharmacotherapy — no. (%)	2687 (34.0)	2664 (33.6)
High-density lipoprotein cholesterol		
Median (IQR) — mg/dl	44.4 (38.6–52.9)	44.8 (38.6–53.7)
<40 mg/dl — no. (%)	2786 (35.2)	2646 (33.4)
Smoker — no. (%)‡	1656 (21.0)	1572 (19.8)
Renal dysfunction — no. (%)§	2374 (30.0)	2410 (30.4)
Polyvascular disease — no. (%)	1187 (15.0)	1185 (15.0)

Number of subjects on Placebo (all subjects)



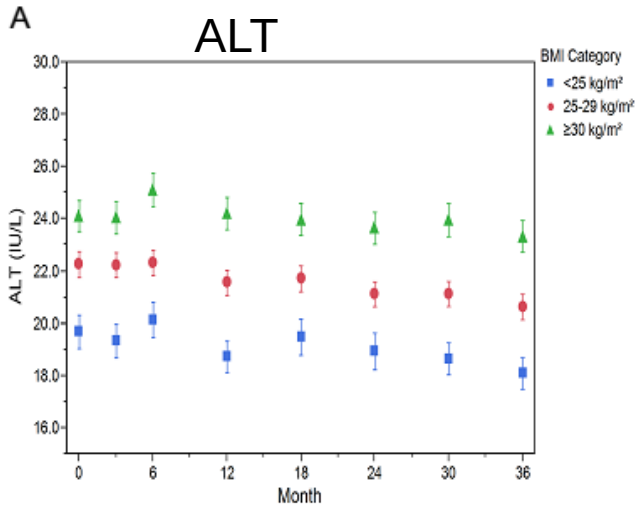
Subject characteristics

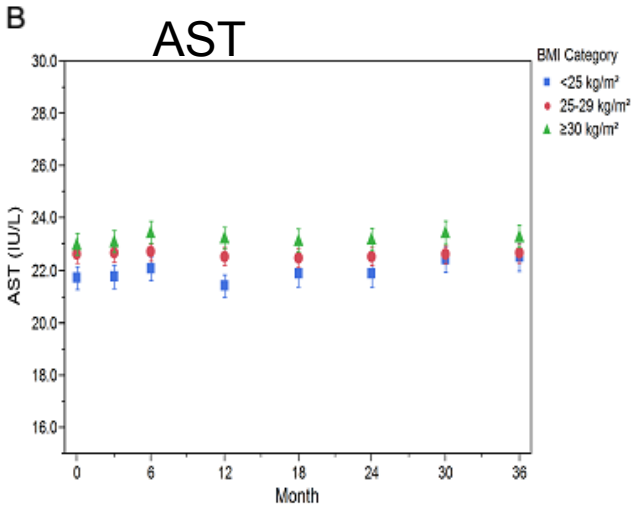
	STABILITY		DIA Trials	
	Baseline	36 months	Baseline	12 Months
	n=4264		n=308	
Age (years), mean (SD)	64.2 (9.1)	67.2 (9.1)	62.4 (7.8)	63.4 (7.8)
BMI (kg/m^2), mean (SD)	28.8 (4.9)	28.9 (5.0)	31.9 (5.6)	31.7 (5.5)
Males, n (%)	3,525 (83)	3,525 (83)	172 (56)	172 (56)
T2D, n (%)	1,605 (38)	1,605 (38)	308 (100)	308 (100)
HbA_{1c} %, mean (SD)	7.3 (1.4)	7.4 (1.5)	7.8 (0.8)	7.5 (0.9)
eGFR <60 mL/ min/1.73m^2, n (%)	566 (13)	611 (14)	86 (28)	89 (29)
Current smoker, n (%)	1,257 (29)	630 (14.8)	26 (8.4)	26 (8.4)

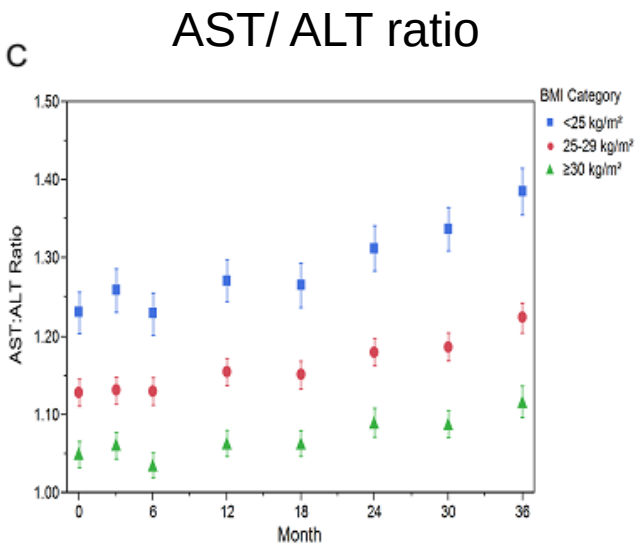
GSK Results

STABILITY: ALT, AST & AST/ALT ratio

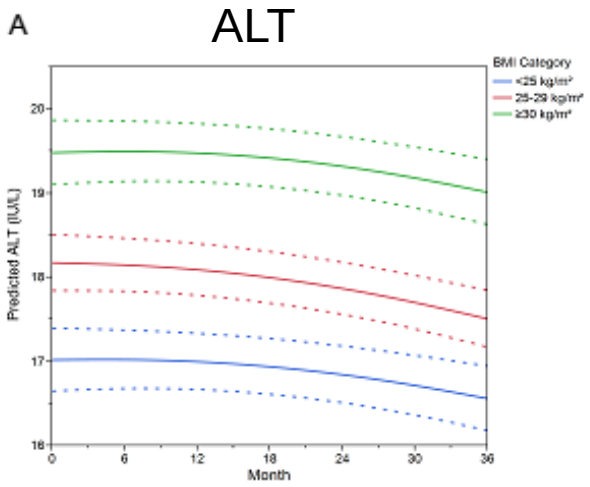
Baseline BMI Category

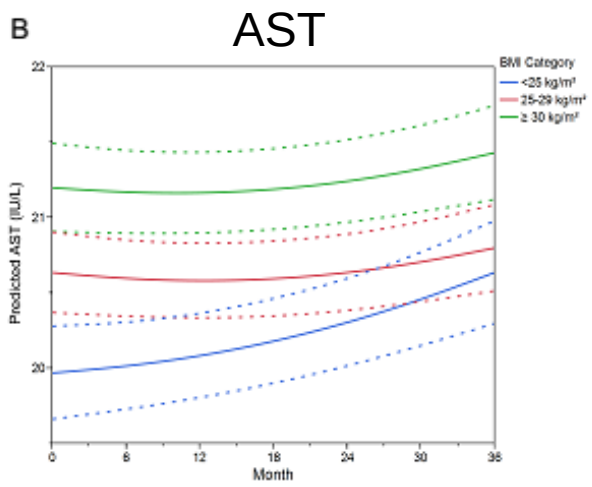


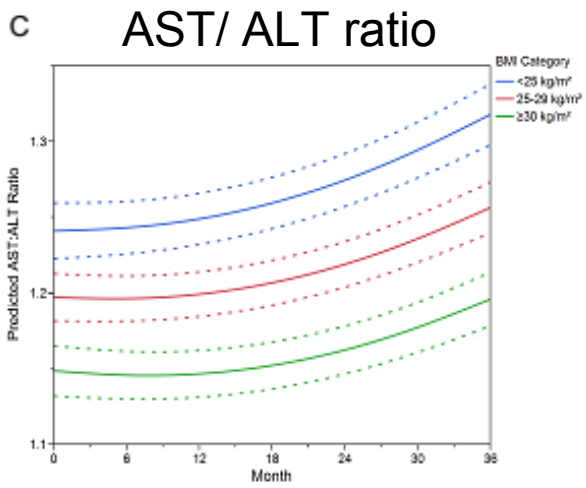




STABILITY: Modeled data of the association of ALT, AST and AST/ALT ratio to visit BMI





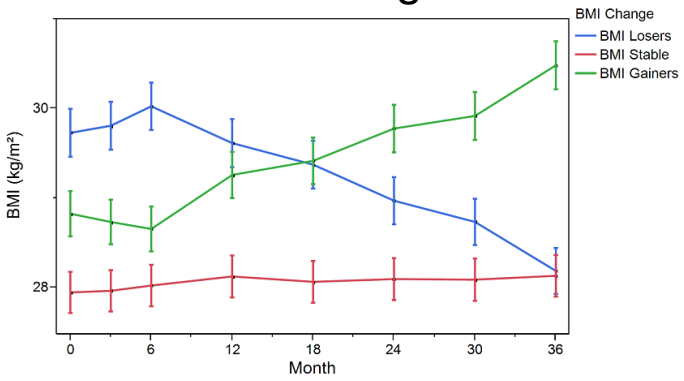


Estimates adjusted for baseline age, gender, smoking, T2D status and eGFR category and incident cardiovascular events

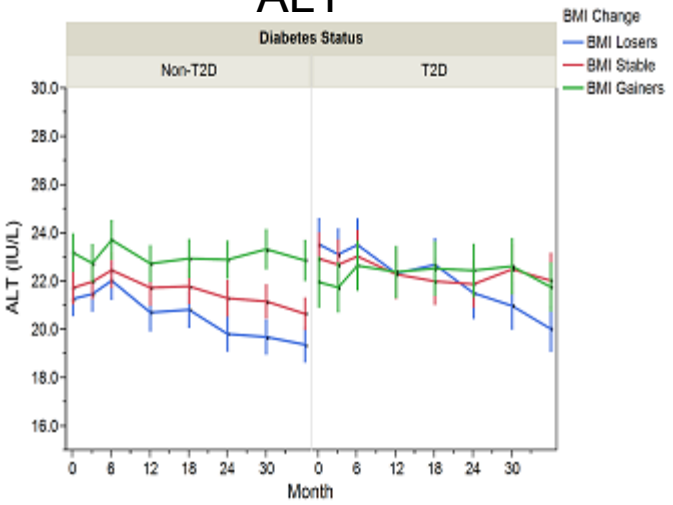
mean ± 95% CI

STABILITY: Effect of Change in BMI *(type 2 diabetes status)*

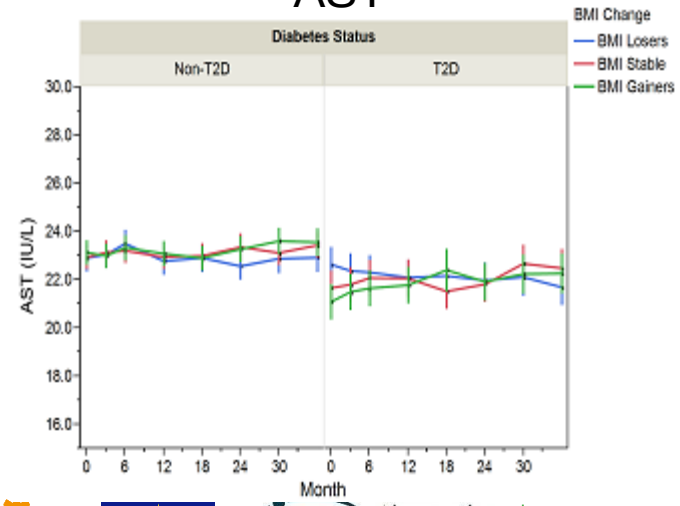
BMI change



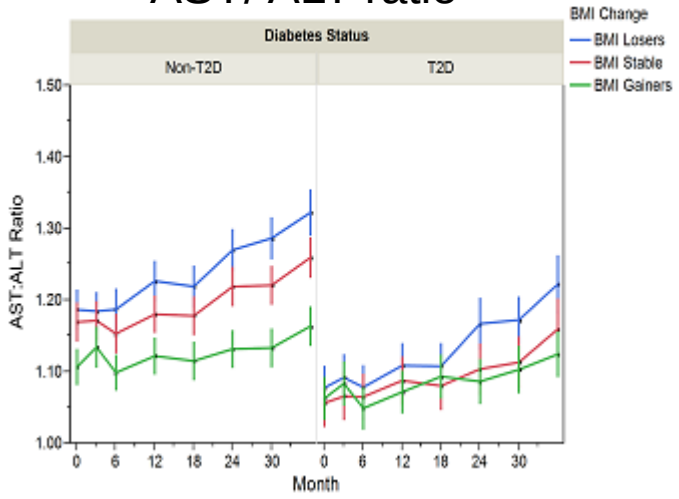
ALT



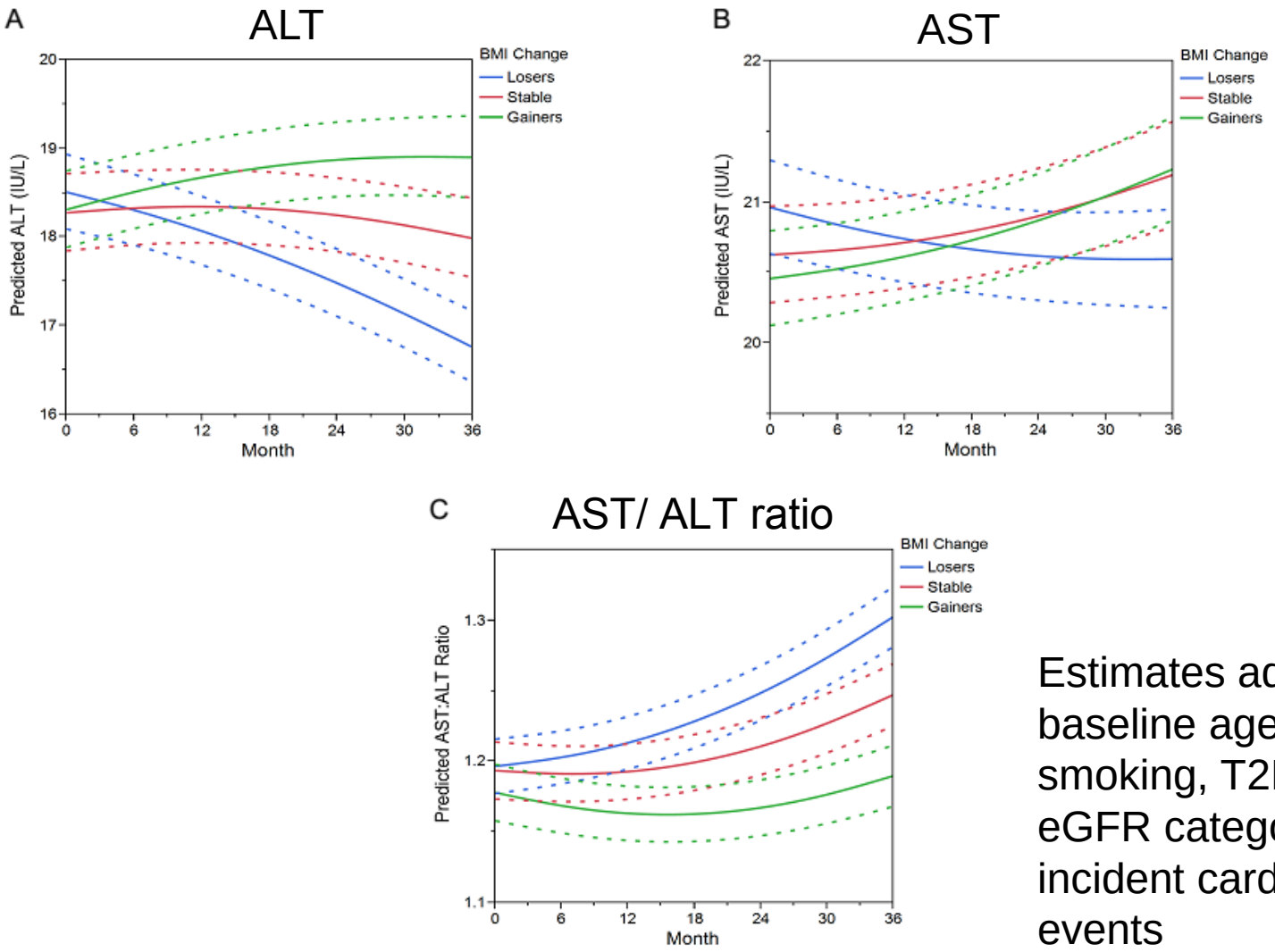
AST



AST/ ALT ratio



STABILITY: Modeled data of the association of ALT, AST and AST/ALT ratio in the 'BMI Gainer', 'BMI Loser' and 'Stable BMI' tertiles



Estimates adjusted for baseline age, gender, smoking, T2D status and eGFR category and incident cardiovascular events

mean ± 95% CI

SUMMARY

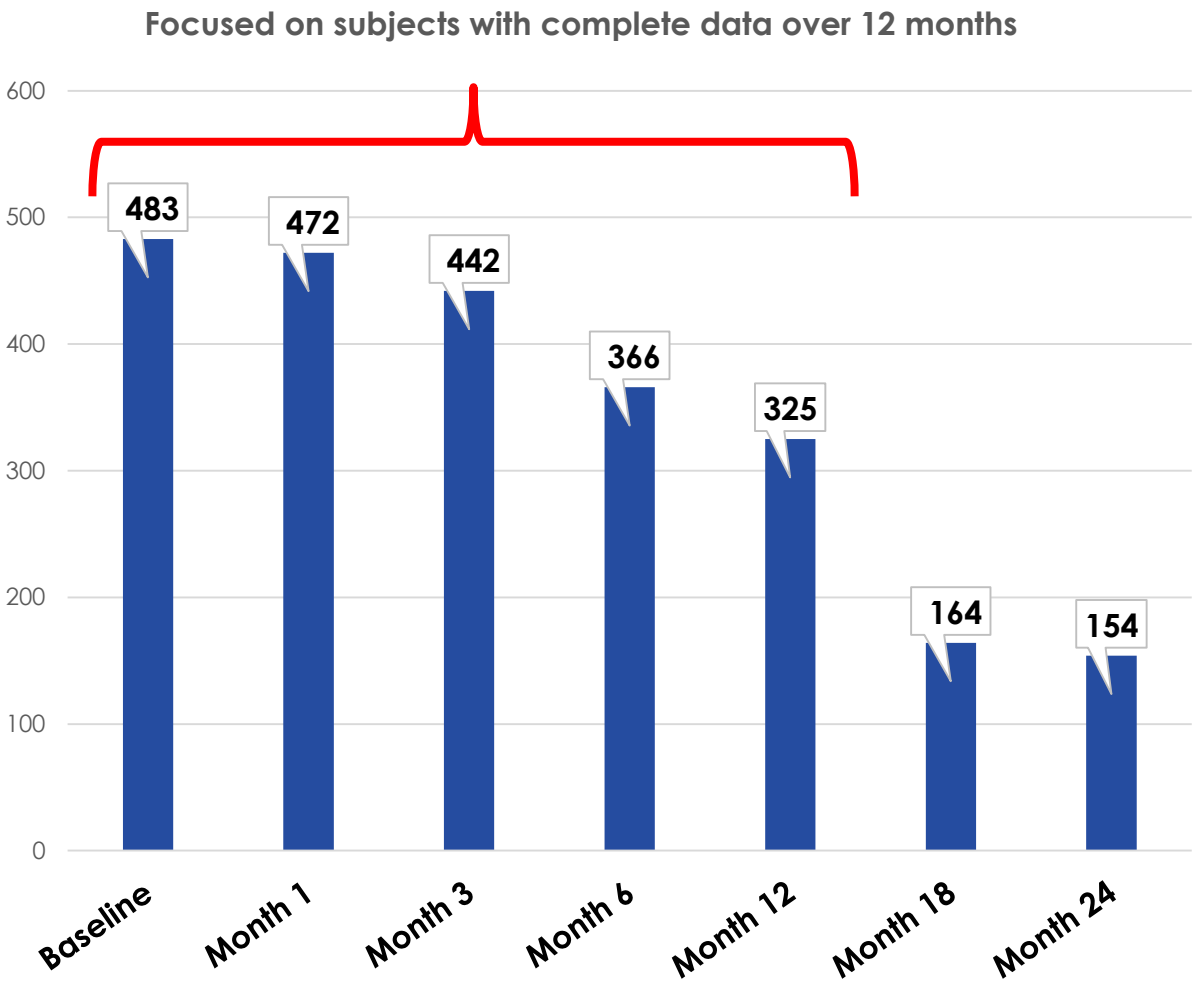
- Clinical trials can be a rich source of longitudinal data for analysis of ‘Natural History’ of a disease or condition
- Need to control for the impact of subject selection criteria and subject drop-outs (important when performing meta-analyses across trials)
- ‘Normalisation’ procedures may need to be implemented for laboratory endpoints to correct for variations in analytical procedures and reference ranges

Janssen Results

Janssen Placebo Data: 3 Completed Phase 3 Trials

Trial	NCTID* (Janssen Identifier)	Phase 3 Clinical Trial Population	Duration	Eligibility Criteria			
				Age, years	HbA1c, %	FPG,mg/dL	eGFR (mL/min/1.73m^2)
1	NCT01106625 (DIA3002)	Subjects with T2DM on metformin and sulphonylurea	52 weeks	18-80	≥7.0 - ≤ 10.5	<270 (15 mmol/L)	≥ 55
2	NCT01064414 (DIA3004)	Subjects with T2DM with moderate renal impairment	52 weeks	≥ 25	≥7.0 - ≤ 10.5	<270 (15 mmol/L at Week-2)	≥30 - <50
3	NCT01106651 (DIA3010)	Older Subjects with T2D	104 weeks	55-80	≥7.0 - ≤ 10.0	<270 (15 mmol/L at Week-2)	≥ 50

Number of Subjects on Placebo in 3 DIA Trials Combined



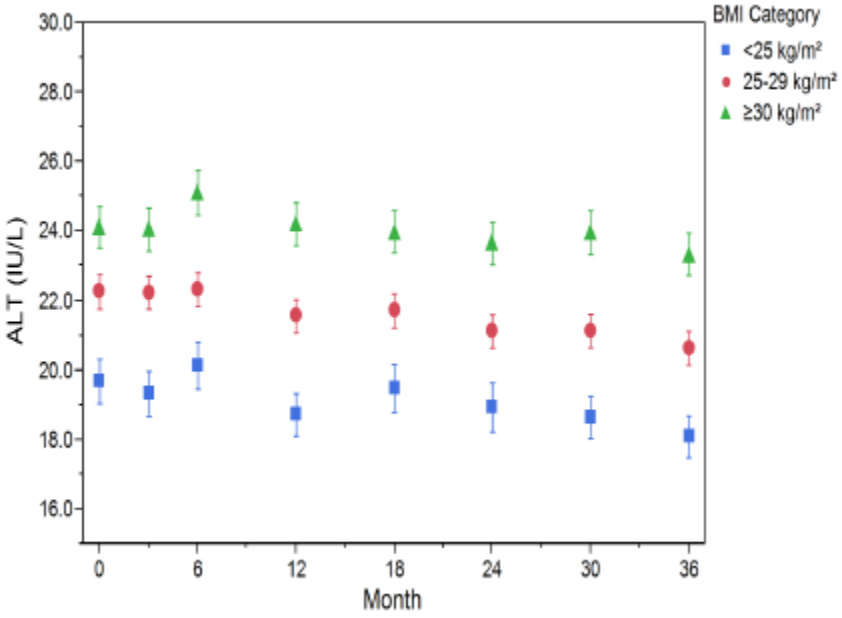
Janssen: Subject Characteristics

	STABILITY		DIA Trials	
	Baseline	36 months	Baseline	12 Months
	n=4264		n=308	
Age (years), mean (SD)	64.2 (9.1)	67.2 (9.1)	62.4 (7.8)	63.4 (7.8)
BMI (kg/m^2), mean (SD)	28.8 (4.9)	28.9 (5.0)	31.9 (5.6)	31.7 (5.5)
Males, n (%)	3,525 (83)	3,525 (83)	172 (56)	172 (56)
T2D, n (%)	1,605 (38)	1,605 (38)	308 (100)	308 (100)
HbA_{1c} %, mean (SD)	7.3 (1.4)	7.4 (1.5)	7.8 (0.8)	7.5 (0.9)
eGFR <60 mL/ min/1.73m^2, n (%)	566 (13)	611 (14)	86 (28)	89 (29)
Current smoker, n (%)	1,257 (29)	630 (14.8)	26 (8.4)	26 (8.4)

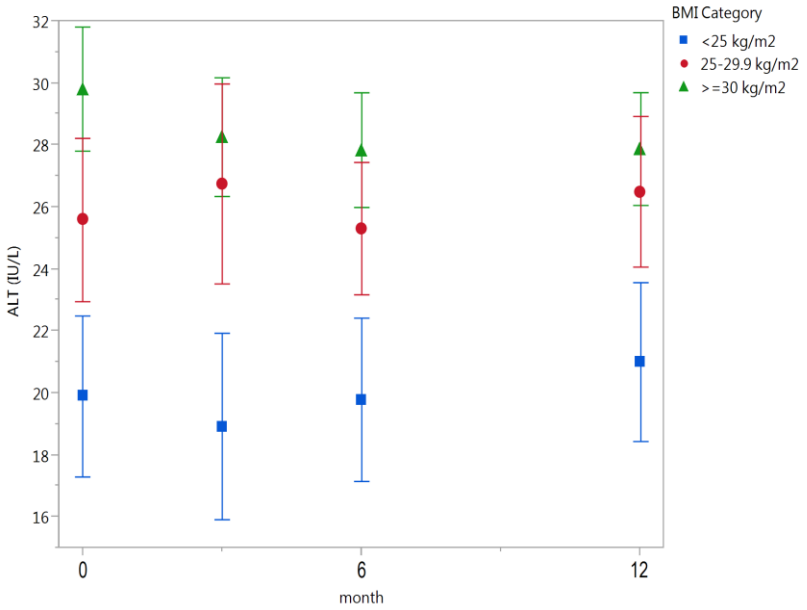
STABILITY versus DIA Trials : ALT

Baseline BMI Category

STABILITY



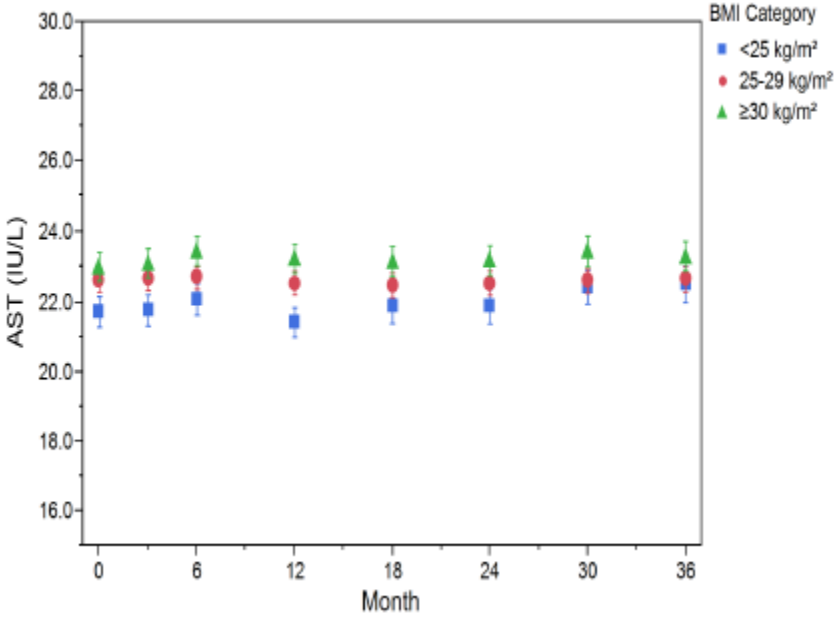
DIA Trials



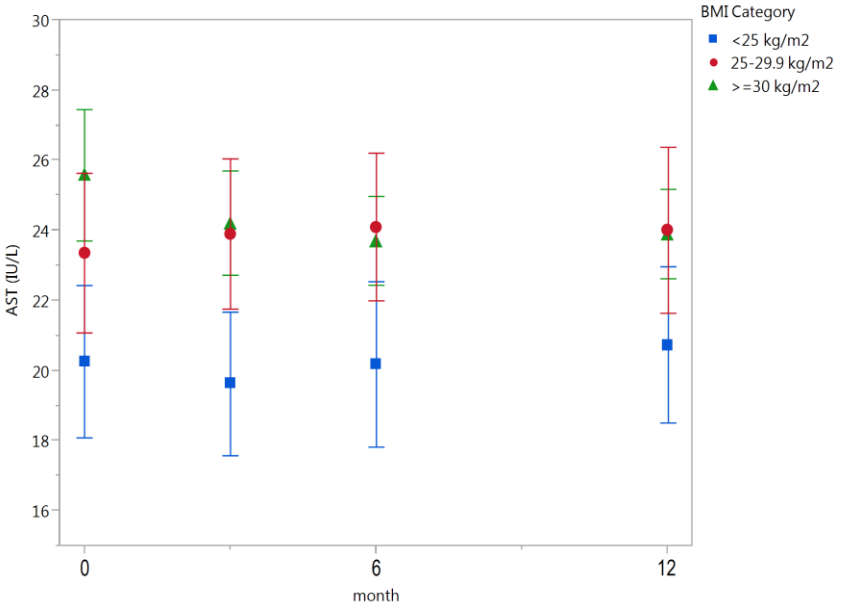
STABILITY versus DIA Trials : AST

Baseline BMI Category

STABILITY



DIA Trials

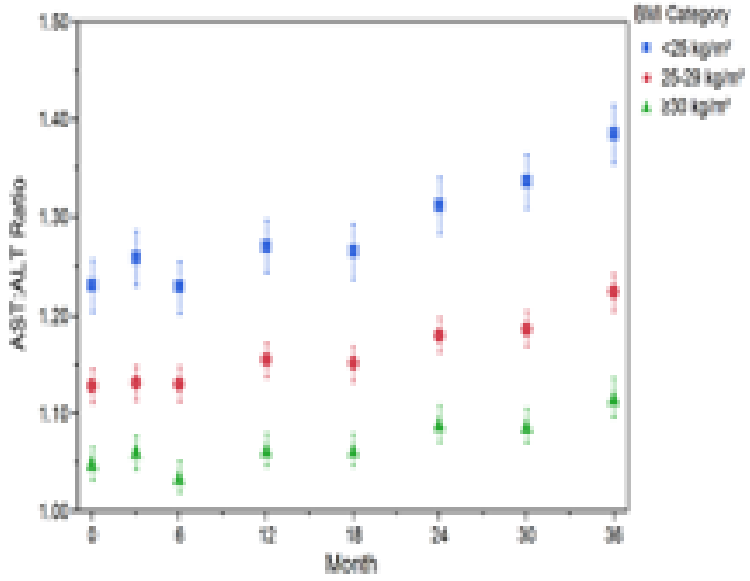


BMI = Body Mass Index

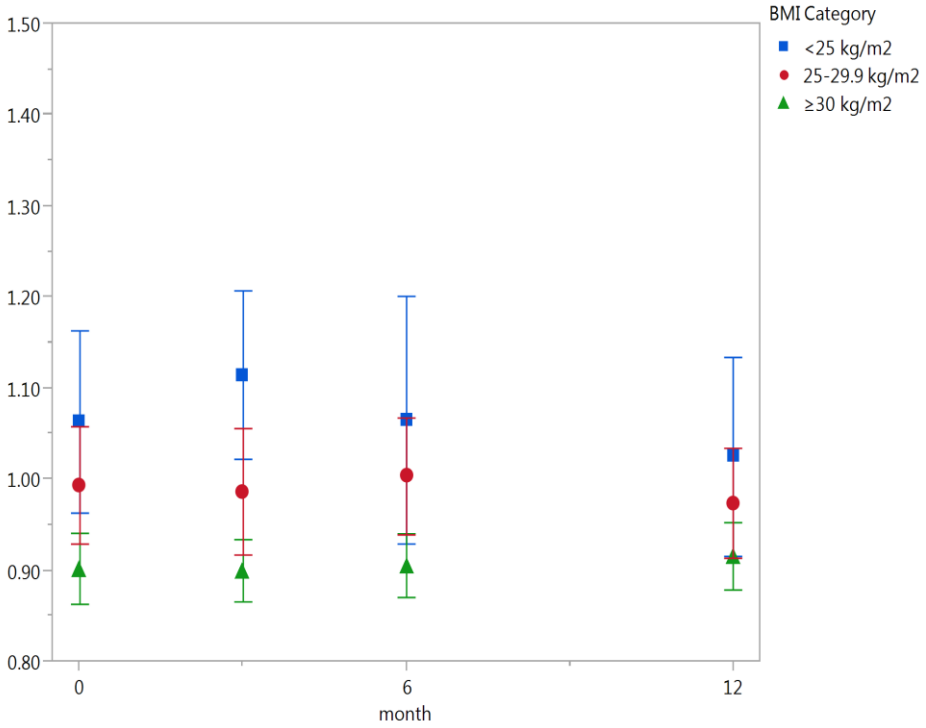
STABILITY versus DIA Trials : AST/ALT Ratio

Baseline BMI Category

STABILITY



DIA Trials

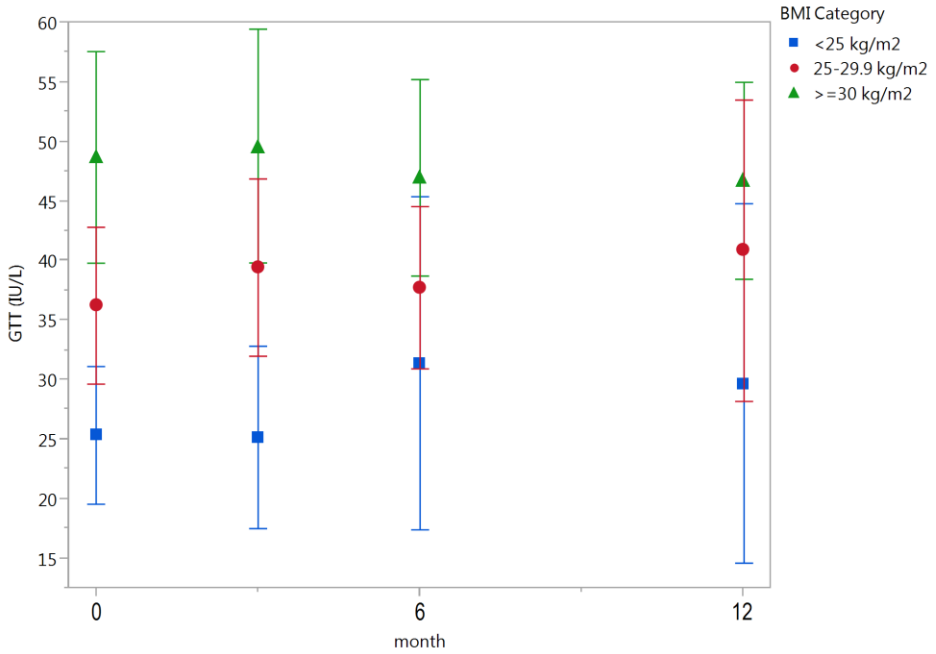


BMI = Body Mass Index

STABILITY versus DIA Trials : GGT
Baseline *BMI Category*

STABILITY

DIA Trials



BMI = Body Mass Index

DIA Trials : Modeled data of the association of ALT, AST and AST/ALT ratio to visit BMI

- Linear mixed models were also applied on 3 the DIA trials corecting for baseline age, gender, smoking and eGFR category but not for T2D status (since no non-T2D subjects in DIA Trials) nor for incident cardiovascular events
- Results are less informative because of :
  - Limited combined longitudinal follow-up time (only 12 months)
  - Relatively limited number of subjects in different sub-groups
  - Wide 95% confidence intervals of estimates

SUMMARY

- Janssen DIA placebo trial data show similar pattern in ALT, GGT, AST and the AST/ALT ratio as noted in the GSK STABILITY trial placebo data

- Precompetitive sharing of data and analyses is feasible

- The use of the use of liver biomarkers in this pilot provided insights that need to be confirmed by amalgamation of further datasets

Other Initiatives of data sharing within Pharma

- Janssen has teamed up with the Yale University Open Data Access (YODA) Project for the responsible sharing of clinical research data to researchers

- TransCelerate has set up the Placebo and Standard of Care (PSoC) Initiative to enable the sharing of de-identified data – from subjects either on placebo or the active ingredient

Acknowledgements

Janssen: Geert Byttebier, Elisa Fabbrini, Gary Meininger, Barry Schwab, and Bart Vannieuwenhuyse

GSK: Myriam Alexander, Nick Galwey, Derek Nunez, Dawn Waterworth and Laura Yerges-Armstrong

University of Glasgow: Naveed Sattar

University of Pisa School of Medicine: Ele Ferrannini

COFFEE BREAK

Panel Discussion

Chair: Nigel Hughes

Janssen Pharma R&D


EMIF


efpia
imi
innovative
medicines
initiative

Closing Statement

Prof Simon Lovestone

Oxford University