



Leveraging Exploratory Data and Analysis for Alzheimer's Disease

DH11

Challenges and Wins of Pooling Internal and External Alzheimer's Data for Secondary Use

Michaela Mertes, Principal Data Sciences Product Leader,
michaela.mertes@roche.com

Iska Gans, Principal Data Sciences Product Leader,
iska.gans@roche.com

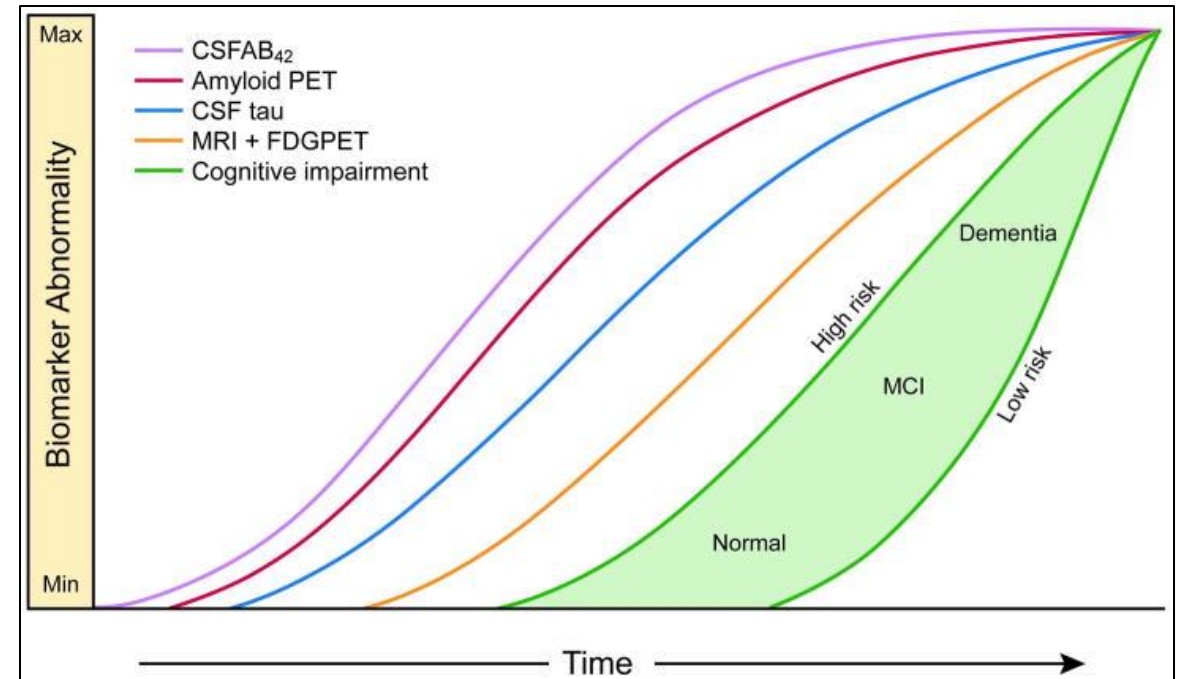
Table of contents

- Information about Alzheimer's Disease
- Fair Data Principles
- Data Flow – Creation of the Data Mart
- Challenges
- Wins

Information about Alzheimer's Disease

What is Alzheimer's Disease?

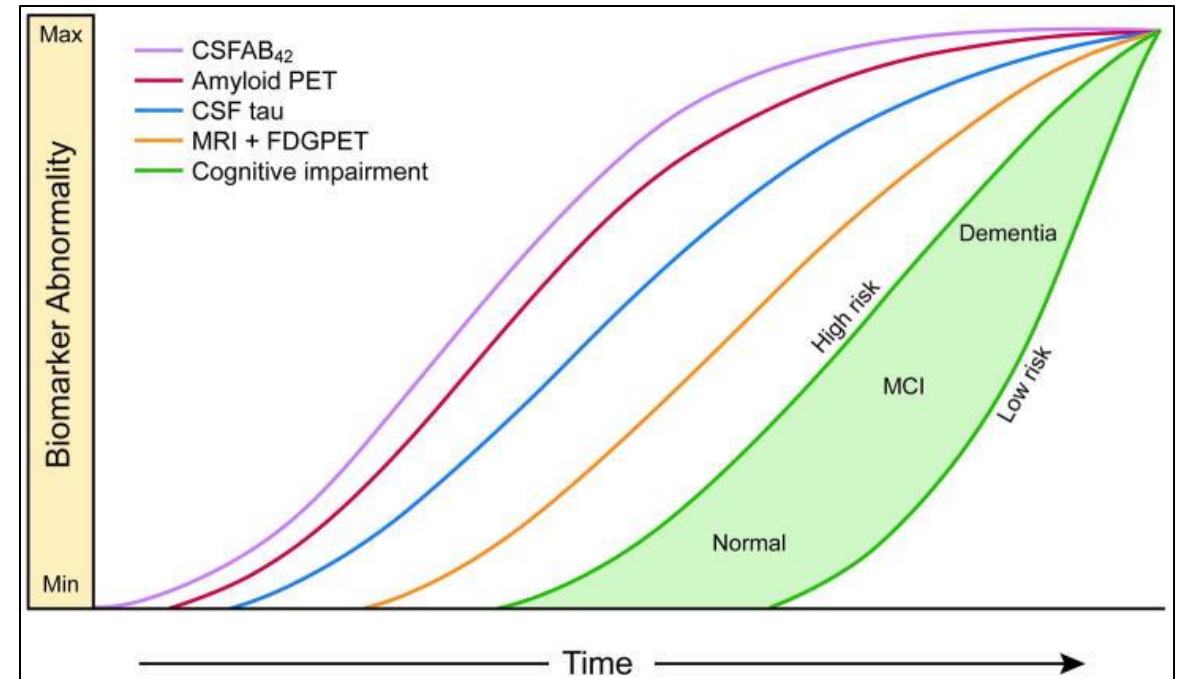
- Alzheimer's is accounting for 60-80% of dementia cases (estimated)



Lancet Neurology vol. 12; Jack CR Jr, et.al, pp207-216;2013

What is Alzheimer's Disease?

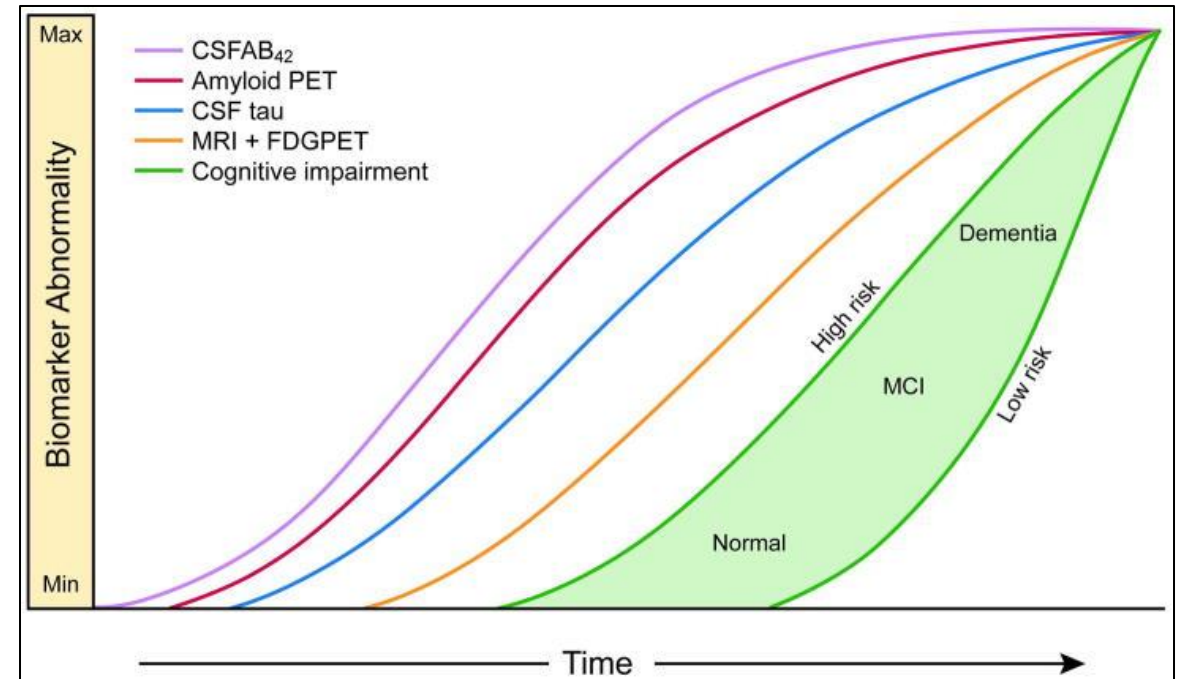
- Alzheimer's is accounting for 60-80% of dementia cases (estimated)
- Worldwide, >55 million people living with Alzheimer's and other dementias



Lancet Neurology vol. 12; Jack CR Jr, et.al, pp207-216;2013

What is Alzheimer's Disease?

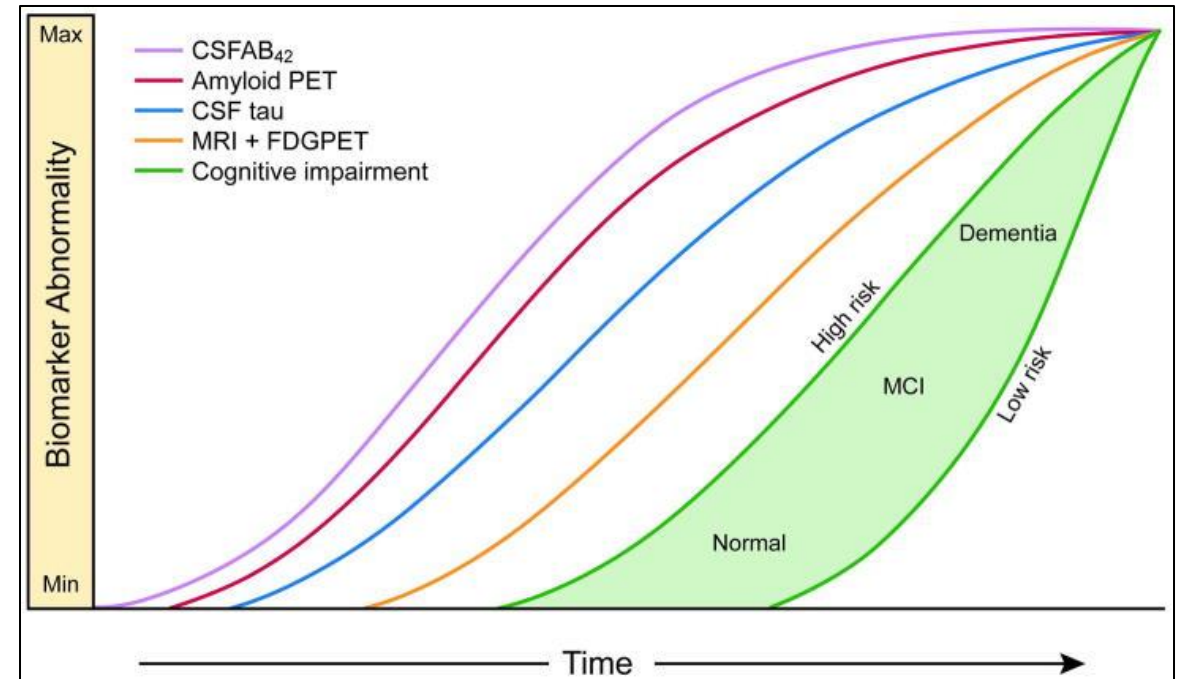
- Alzheimer's is accounting for 60-80% of dementia cases (estimated)
- Worldwide, >55 million people living with Alzheimer's and other dementias
- Caused by damage of nerve cells in the brain



Lancet Neurology vol. 12; Jack CR Jr, et.al, pp207-216;2013

What is Alzheimer's Disease?

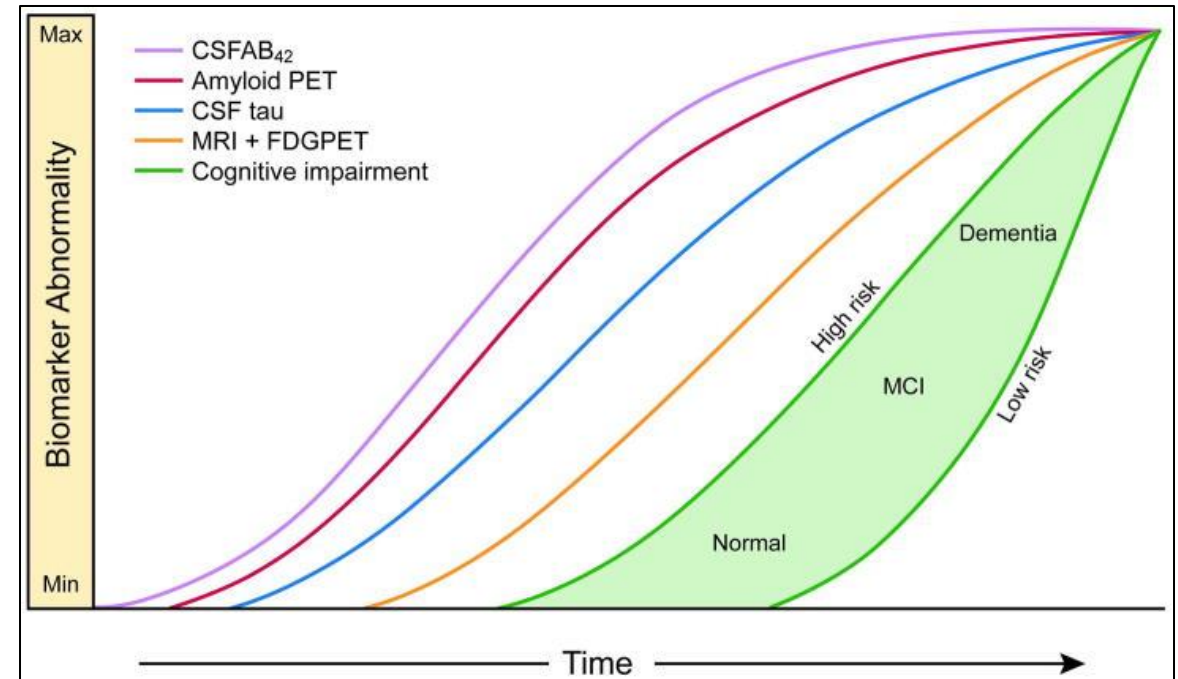
- Alzheimer's is accounting for 60-80% of dementia cases (estimated)
- Worldwide, >55 million people living with Alzheimer's and other dementias
- Caused by damage of nerve cells in the brain
- Irreversible, progressive brain disorder that slowly destroys memory and thinking skills



Lancet Neurology vol. 12; Jack CR Jr, et.al, pp207-216;2013

What is Alzheimer's Disease?

- Alzheimer's is accounting for 60-80% of dementia cases (estimated)
- Worldwide, >55 million people living with Alzheimer's and other dementias
- Caused by damage of nerve cells in the brain
- Irreversible, progressive brain disorder that slowly destroys memory and thinking skills
- Alzheimer's disease begins 20 years or more before memory loss and other symptoms develop.

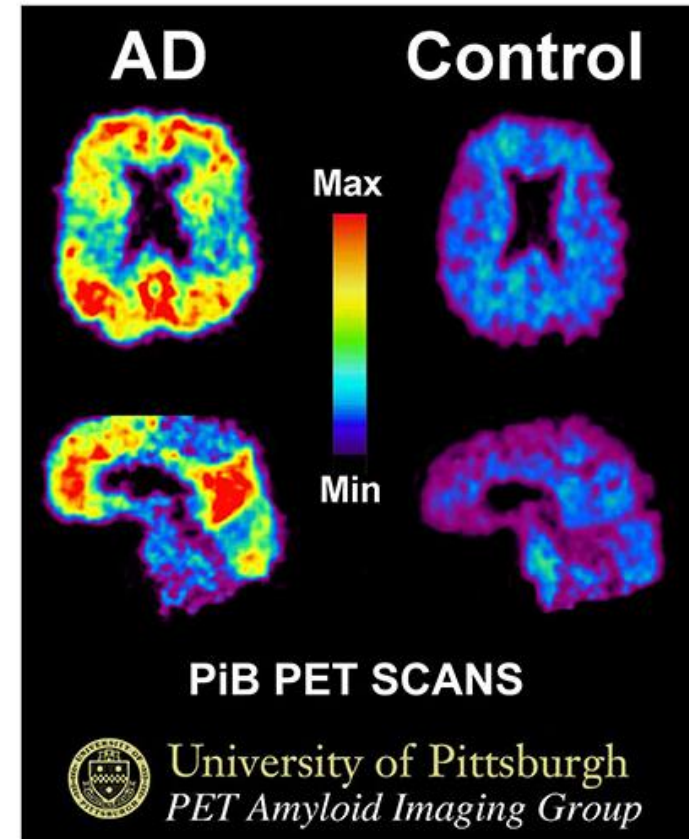


Lancet Neurology vol. 12; Jack CR Jr, et.al, pp207-216;2013

Brain Changes of Alzheimer's Disease

- **Beta-Amyloid Plaques**

Accumulation of the protein fragment beta-amyloid into clumps outside neurons



Pittsburgh compound B (PiB) is currently being studied. The PET brain scan on the left is of a person with mild Alzheimer's disease, and the scan on the right is from a healthy adult. Red areas show the greatest levels of amyloid, and the dark blue indicates no amyloid. The other colors — in order like the colors of a rainbow — show progressively lower levels of amyloid.

Image Credit:
University of Pittsburgh Medical Center

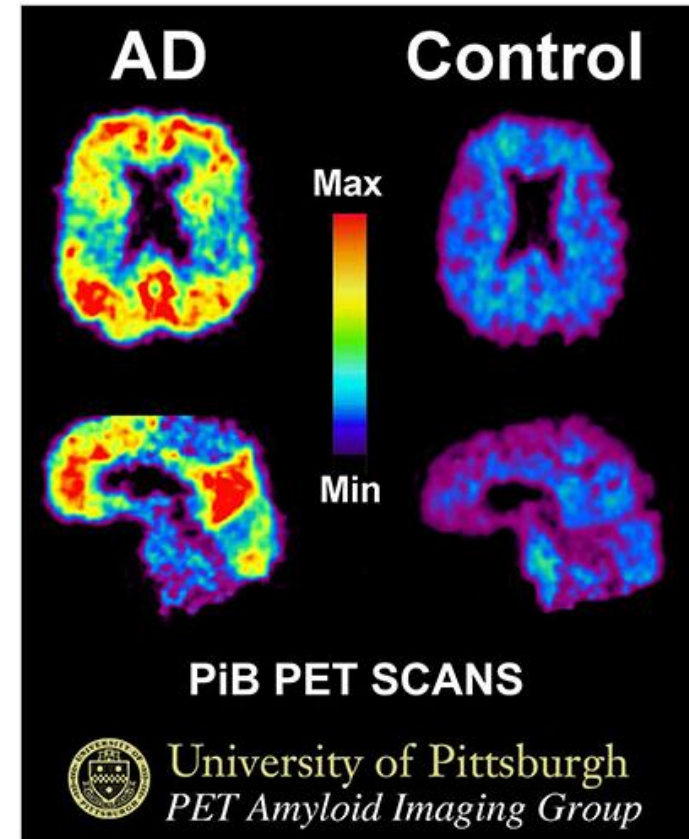
Brain Changes of Alzheimer's Disease

- **Beta-Amyloid Plaques**

Accumulation of the protein fragment beta-amyloid into clumps outside neurons

- **Tau Tangles**

Accumulation of an abnormal form of the protein tau (called tau tangles) inside neurons.



Pittsburgh compound B (PiB) is currently being studied. The PET brain scan on the left is of a person with mild Alzheimer's disease, and the scan on the right is from a healthy adult. Red areas show the greatest levels of amyloid, and the dark blue indicates no amyloid. The other colors — in order like the colors of a rainbow — show progressively lower levels of amyloid.

Image Credit:
University of Pittsburgh Medical Center

Brain Changes of Alzheimer's Disease

- **Beta-Amyloid Plaques**

Accumulation of the protein fragment beta-amyloid into clumps outside neurons

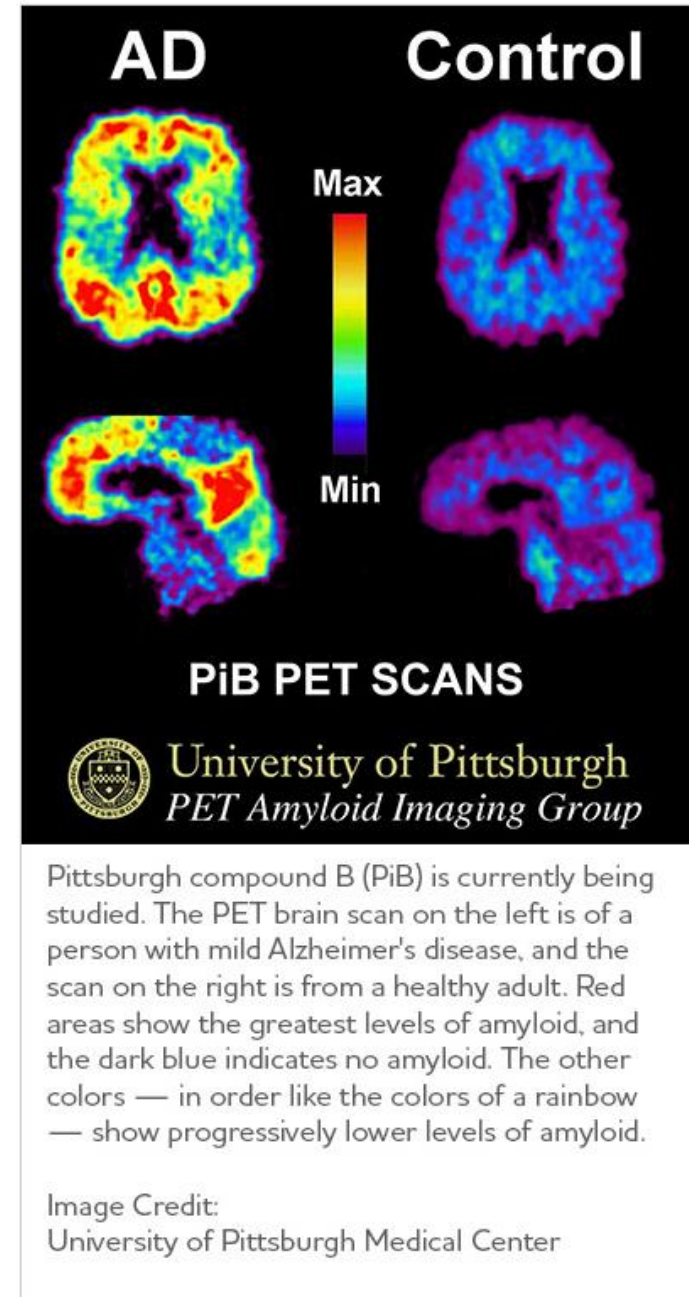
- **Tau Tangles**

Accumulation of an abnormal form of the protein tau (called tau tangles) inside neurons.

- **Neurodegeneration**

Destruction of neurons

[2024 Alzheimer's Disease Facts and Figures
https://en.wikipedia.org/wiki/Pittsburgh_compound_B](https://en.wikipedia.org/wiki/Pittsburgh_compound_B)



Brain Changes of Alzheimer's Disease

- **Beta-Amyloid Plaques**

Accumulation of the protein fragment beta-amyloid into clumps outside neurons

- **Tau Tangles**

Accumulation of an abnormal form of the protein tau (called tau tangles) inside neurons.

- **Neurodegeneration**

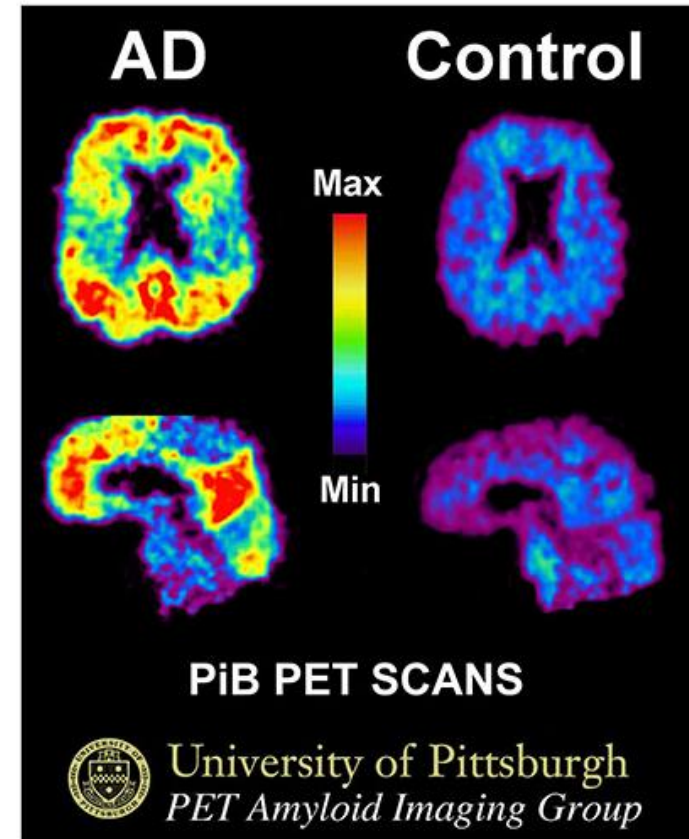
Destruction of neurons

- **Atrophy**

Decreased brain volume

[2024 Alzheimer's Disease Facts and Figures](https://en.wikipedia.org/wiki/Pittsburgh_compound_B)

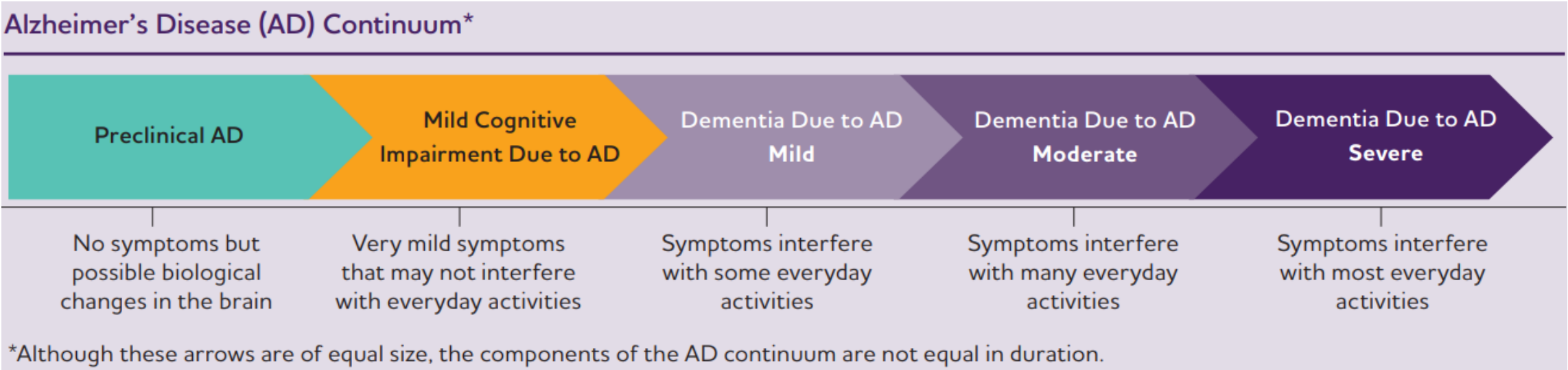
https://en.wikipedia.org/wiki/Pittsburgh_compound_B



Pittsburgh compound B (PiB) is currently being studied. The PET brain scan on the left is of a person with mild Alzheimer's disease, and the scan on the right is from a healthy adult. Red areas show the greatest levels of amyloid, and the dark blue indicates no amyloid. The other colors — in order like the colors of a rainbow — show progressively lower levels of amyloid.

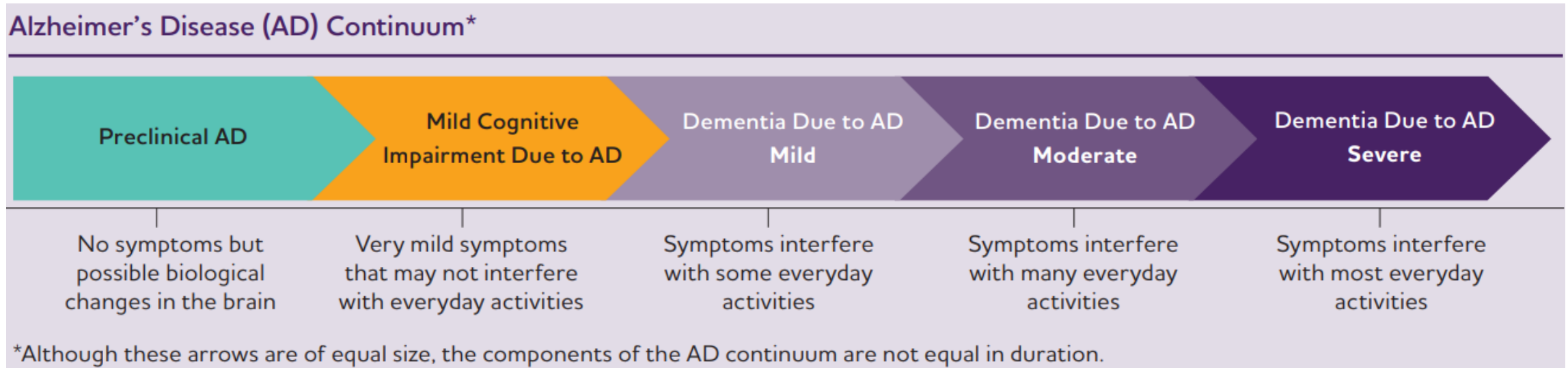
Image Credit:
University of Pittsburgh Medical Center

Alzheimer's Disease (AD) Continuum



[2024 Alzheimer's Disease Facts and Figures](#)

Alzheimer's Disease (AD) Continuum



[2024 Alzheimer's Disease Facts and Figures](#)

Alzheimer's data mart includes demographics, clinical data, images, fluid biomarkers, genetics)
up to 25 years follow-up data in all stages

Started May 2019

LeAD Data Mart

Leverage Exploratory Analysis and Data in AD

- 35 studies
(14 in-house,
21 external studies)
- >75,000 patients

R&D Dataset Portal

[Home](#) / [Catalogs](#) / [Alzheimer's Disease](#)

ALZHEIMER'S

Alzheimer's Disease

Information on PROJECT, DATA FLOW, and AVAILABLE DATASETS Leveraging exploratory Analysis and Data for Alzheimer's Disease (LeAD) by using the wealth of internal and external... [read more](#)

[Datasets](#)
[Activity Stream](#)
[About](#)

Information on PROJECT, DATA FLOW, and AVAILABLE DATASETS

Leveraging exploratory Analysis and Data for Alzheimer's Disease (LeAD)

by using the wealth of internal and external data to improve drug development and diagnostics in AD for our molecules (higher PTS, smaller, faster trials).

Overview of curated and integrated variables across all studies is [HERE](#) and more details [HERE](#)

DATA MART INFORMATION
 Alzheimer's disease use case

[Information & Links](#)
[R&D Enabled Insights](#)
[Data Access](#)
[Main Contacts](#)

Description

STUDIES
35

SUBJECTS
76,392

CLINICAL DOMAINS
12

HIGH-DIMENSIONAL BIOMARKER
0

Title	Alzheimer's disease use case
Therapeutic Area	Neuroscience
Lifecycle	IN ACTIVE DEVELOPMENT
Published Release	Release 4
Short Description	Wealth source of Alzheimer Disease patient data from Roche internal studies and external data sources.
Description	The Alzheimer's disease use case (also known as "LeAD - Leveraging Exploratory Analysis and Data for Alzheimer's Disease") team mission is to leverage the wealth of Roche internal and external data to improve how we develop drugs and diagnostics in AD for our current and future molecules (higher PTS, smaller, faster trials). The project team was established in May 2019 to continue and expand the work that MIND4AD has started. The MIND4AD and LeAD Use Cases and data mart are sponsored by the Neuro BTALT (Biometrics TA Leadership Team).

Lead Data Mart

Descriptive Statistics

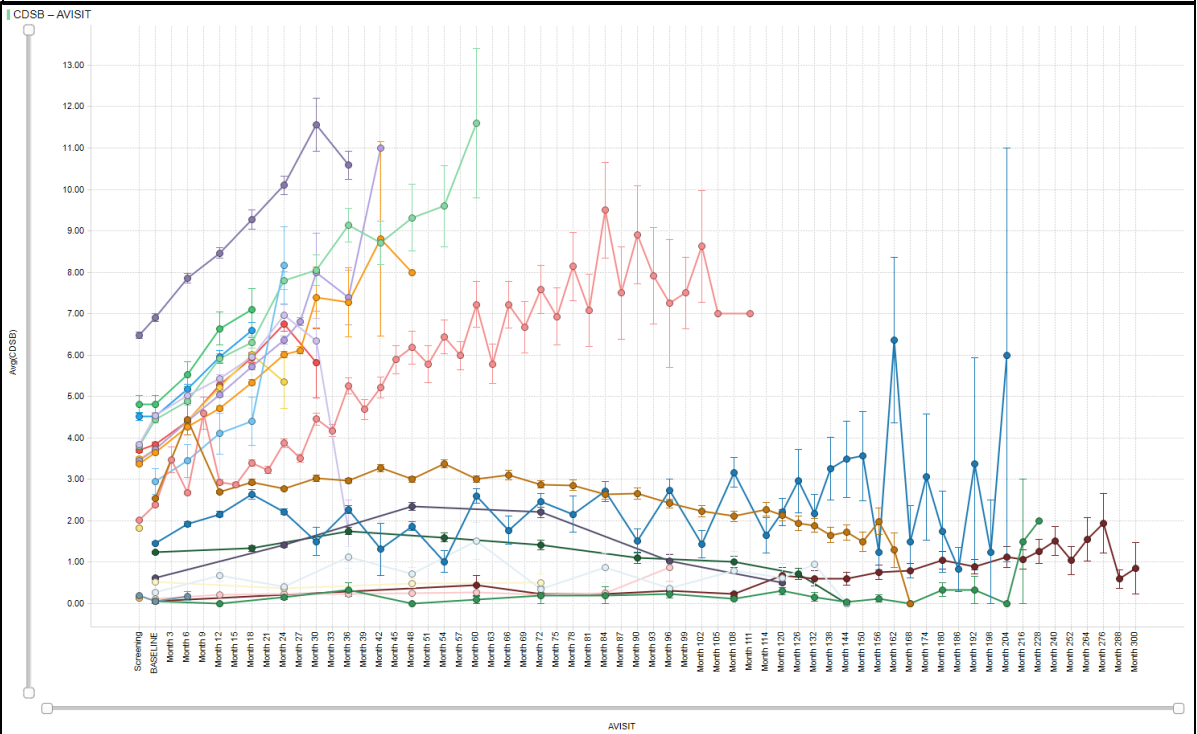
Summary of Baseline/Screening Characteristics - Categorical Variables (First Visit)

Category	Value	n
SEX	Subtotal	76348
		31
	F	44124
	M	30707
RACE	Subtotal	
	AMERICAN INDIAN OR ALA S...	
	ASIAN	
	BLACK OR AFRICAN AMERIC...	
	MULTIPLE	
	NATIVE HAWAIIAN OR OTHE...	
	OTHER	
	UNKNOWN	
	WHITE	

Summary of Baseline Characteristics - Categorical Variables (First Visit)

Category	(Column Names)
AGE	n
	Mean
	StdDev
	Median
	Min
	Max
BBMI	n
	Mean
	StdDev
	Median
	Min
	Max
BHGHTCM	n
	Mean
	StdDev
	Median
	Min
	Max

Clinical Dementia Rating Scale (CDR)
Average of Sum of Boxes – per Study by analysis visit



FAIR Data Shared



Findable
"I know where
data is"



Accessible
"I can access
data I need"



Interoperable
"I use consistent
standards & terms"



Reusable
"I use existing data to
answer new questions"

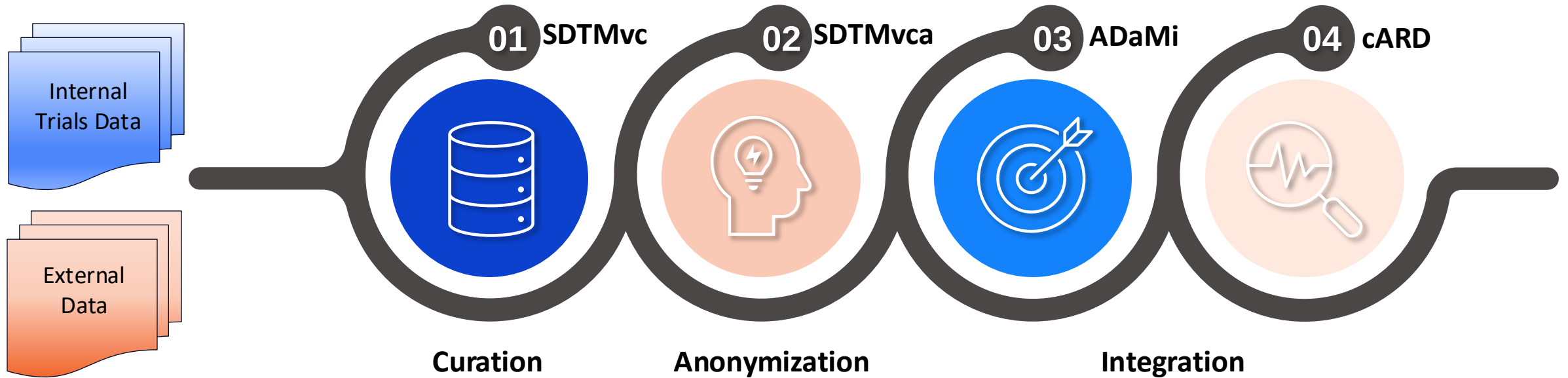
Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process.

Wilkinson M. et al. The FAIR Guiding Principles for scientific management and stewardship (2016)



Data flow – Creating the LeAD data mart

Data Flow – Data Models



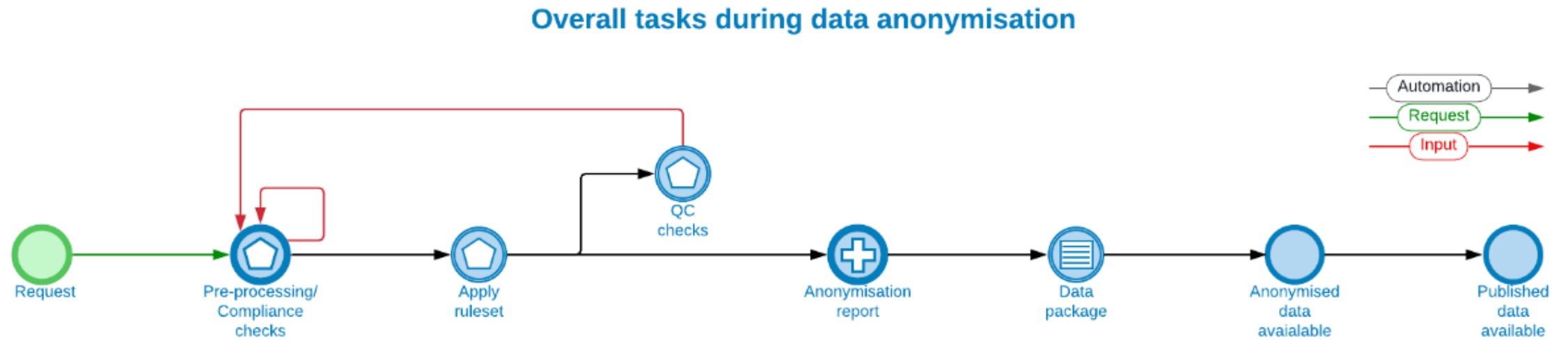
Curation Process



Anonymization

Anonymization is the overall process of turning data about an identifiable person ('personal data') into data which no longer relates to an identifiable person, thereby reducing the risk of re-identification to a negligible level.

Process Flow when using the Anonymization Service



Challenges

Challenges in curation

- Combining diverse data sources, formats and varying standards

Challenges in curation

- Combining diverse data sources, formats and varying standards
- Data standardization and harmonization across multiple studies

Challenges in curation

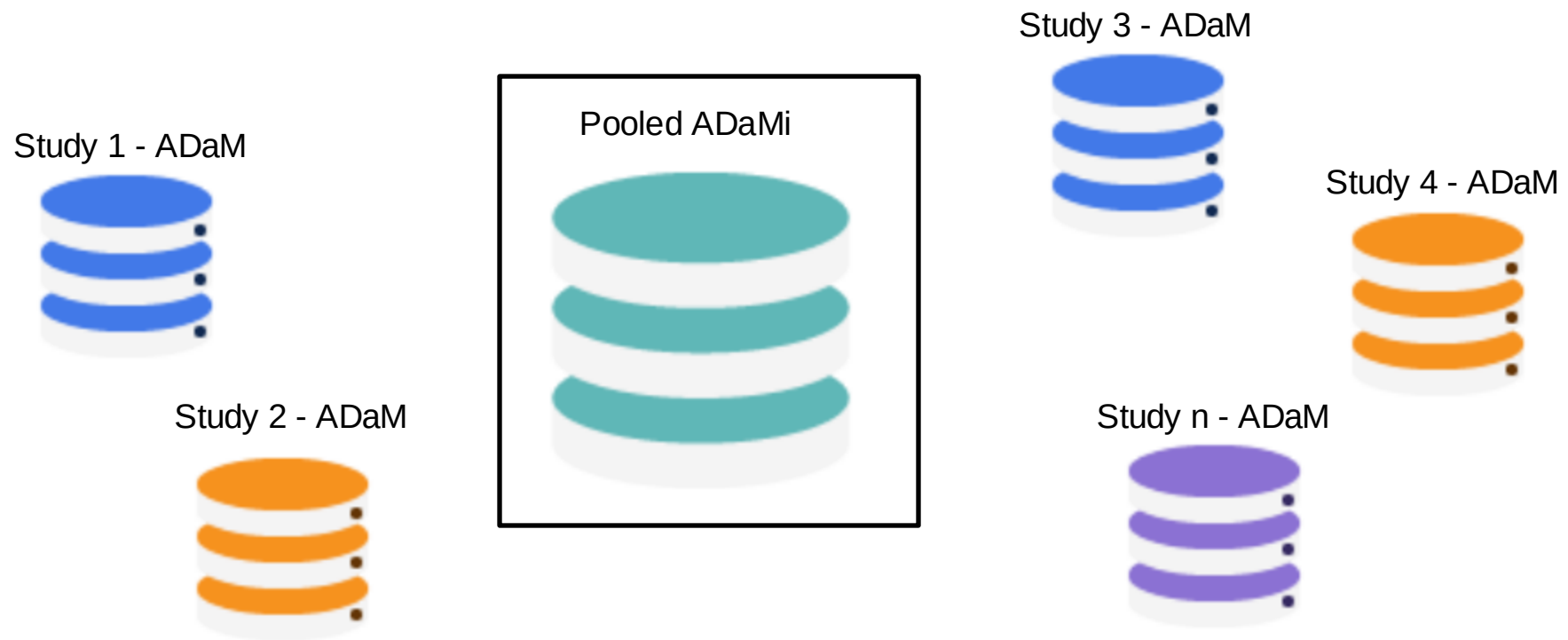
- Combining diverse data sources, formats and varying standards
- Data standardization and harmonization across multiple studies
- Context and metadata around the data is important (e.g. documentation, etc.)

Challenges in curation

- Combining diverse data sources, formats and varying standards
- Data standardization and harmonization across multiple studies
- Context and metadata around the data is important (e.g. documentation, etc.)
- Ensuring interoperability across sources

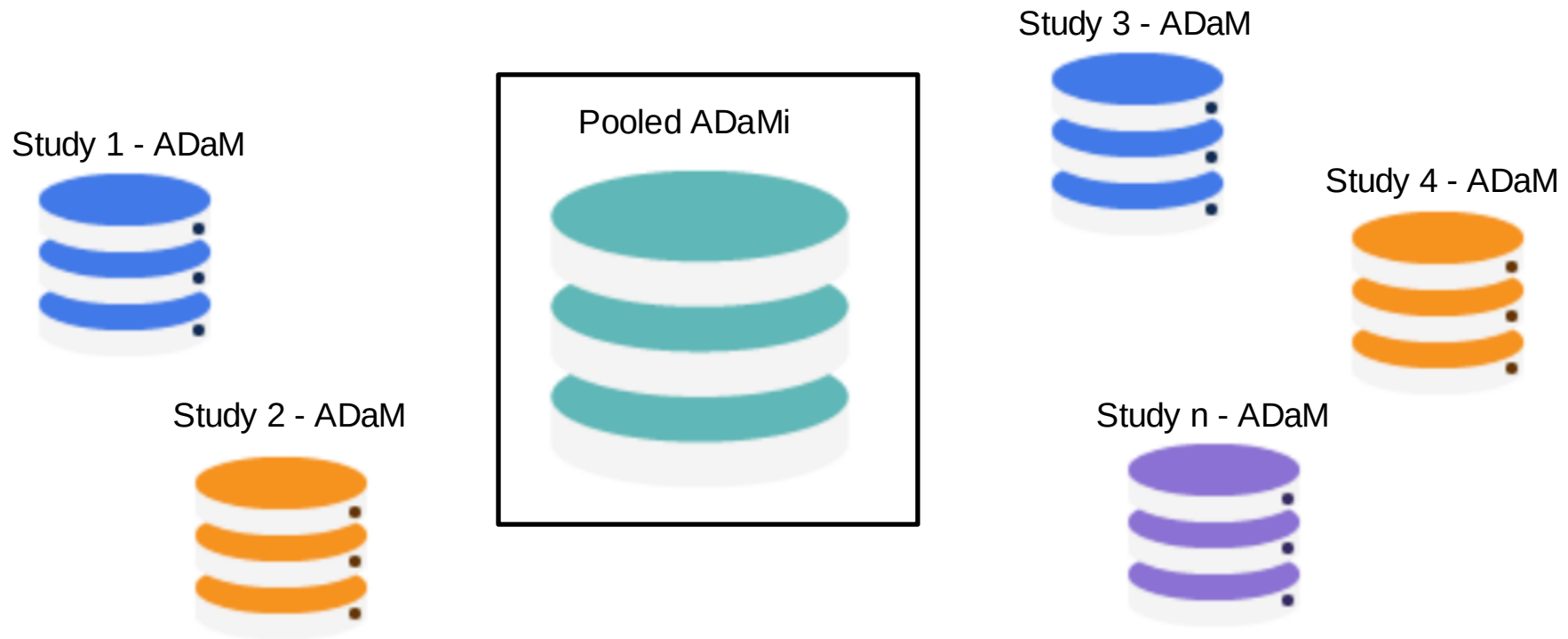
Differences between Study ADaM and pooled ADaMi

- Same definition of Baseline for all studies in ADaMi



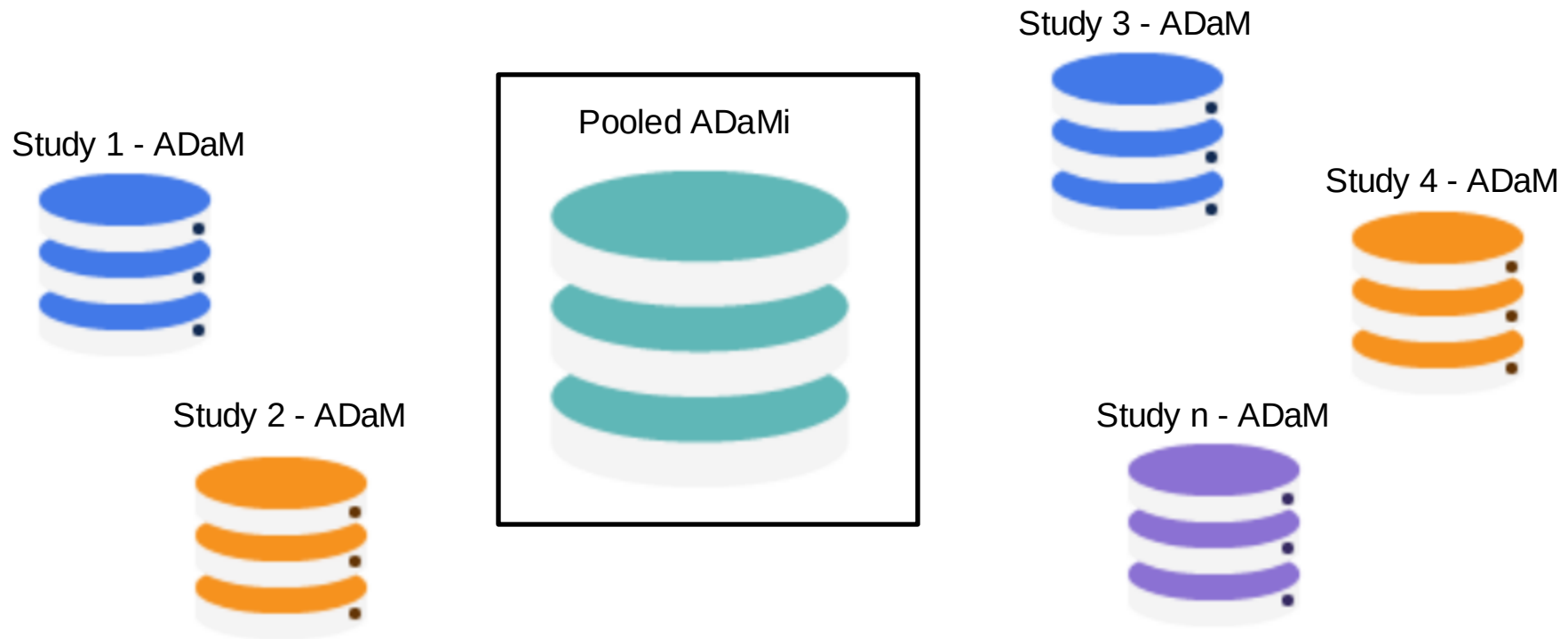
Differences between Study ADaM and pooled ADaMi

- Same definition of Baseline for all studies in ADaMi
- Visit windowing more general



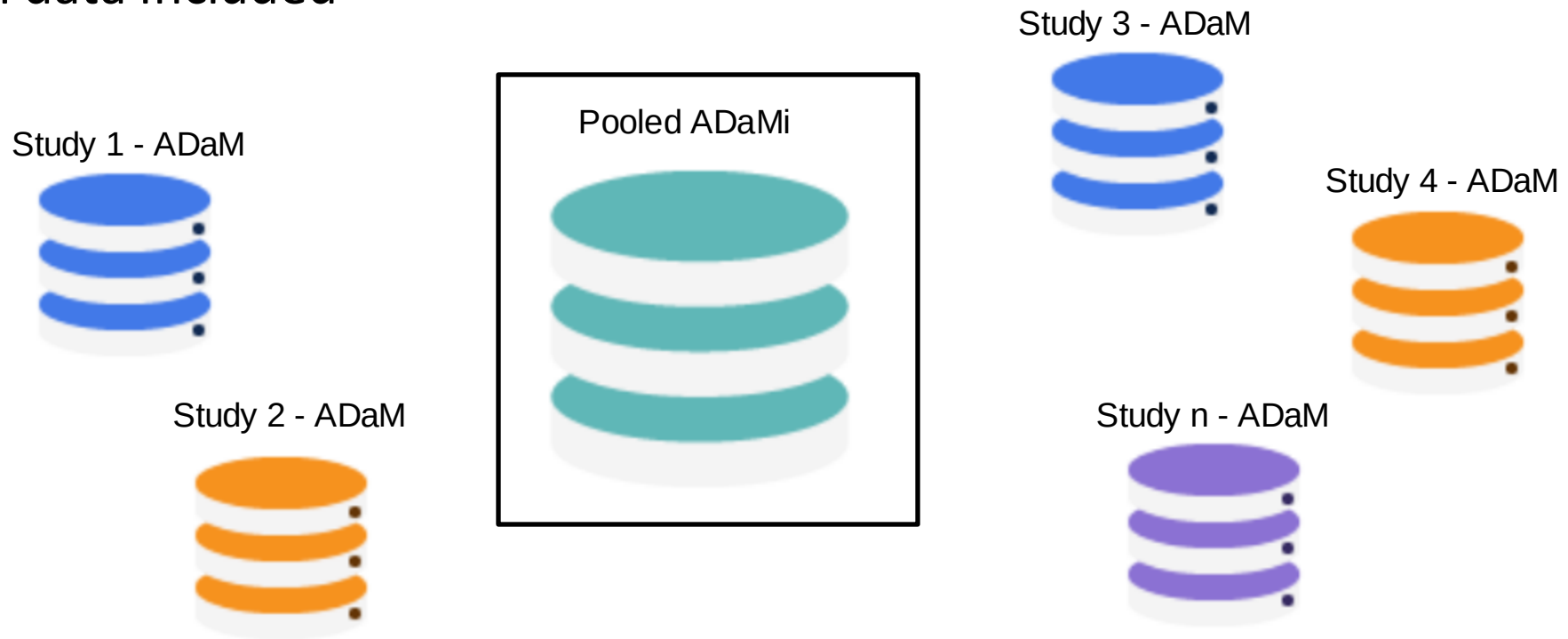
Differences between Study ADaM and pooled ADaMi

- Same definition of Baseline for all studies in ADaMi
- Visit windowing more general
- Screen Failures



Differences between Study ADaM and pooled ADaMi

- Same definition of Baseline for all studies in ADaMi
- Visit windowing more general
- Screen Failures
- Additional data included



Comparison of values, e.g. Laboratory values

Difference between single study and pooling

[SDTM.LB]: CSF Variable [LBTESTCD] = «**AMYLB40**»

PARAM	PARAMCD	LBTESTCD	LBMTDTL	LBMETHOD	LBSMPLID	LBRLOT
CSF: Amyloid Peptid Beta 40 (Elecsys Batch Neurotoolkit) (ng/mL)	CFAB40	AMYLB40	ELECSYS ROCHE DIAGNOSTICS NEUROTOOLKIT	ELECTROCHEMILUMINESCENCE IMMUNOASSAY		
CSF: Amyloid Peptid Beta 40 (Elecsys at Blennow lab Batch at Screening) (ng/mL)	CFAB40E3	AMYLB40	ELECSYS ROCHE DIAGNOSTICS NEUROTOOLKIT	ELECTROCHEMILUMINESCENCE IMMUNOASSAY	ALIQUOT 1	
CSF: Amyloid Peptid Beta 40 (Elecsys at Blennow lab Batch only patients with follow-up) (ng/mL)	CFAB40E4	AMYLB40	ELECSYS ROCHE DIAGNOSTICS NEUROTOOLKIT	ELECTROCHEMILUMINESCENCE IMMUNOASSAY	ALIQUOT 2	
CSF: Amyloid Peptid Beta 40 (Lumipulse) (ng/mL)	CFAB40L	AMYLB40	LUMIPULSE FUJIREBIO	CHEMILUMINESCENT ENZYME		
CSF: Amyloid Peptid Beta 40 (Elecsys Batch Neurotoolkit) (ng/mL) (Lot K05)	CFAB40L1	AMYLB40	ELECSYS ROCHE DIAGNOSTICS NEUROTOOLKIT	ELECTROCHEMILUMINESCENCE IMMUNOASSAY		K05
CSF: Amyloid Peptid Beta 40 (Euroimmune) (ng/mL)	CFAB40M	AMYLB40	EUROIMMUNE	ELISA		
CSF: Amyloid Peptid Beta 40 (Eli Lilly Proprietary Assays) (ng/mL)	CFAB40P	AMYLB40	ELI LILLY PROPRIETARY ASSAYS	ELISA		
CSF: Amyloid Peptid Beta 40 (Elecsys Batch Neurotoolkit) (ng/mL) (Previous Lot)	CFAB40P1	AMYLB40	ELECSYS ROCHE DIAGNOSTICS NEUROTOOLKIT	ELECTROCHEMILUMINESCENCE IMMUNOASSAY		PREVIOUS
CSF: Amyloid Peptid Beta 40 (LCMS at PPD Batch) (ng/mL)	CFAB40PB	AMYLB40	INNOTEST FUJIREBIO	ELISA		

Comparison of values, e.g. Laboratory

Difference between single study and pooling

[SDTM.LB]: CSF Variable [LBTESTCD] = «**AMYL B40**»

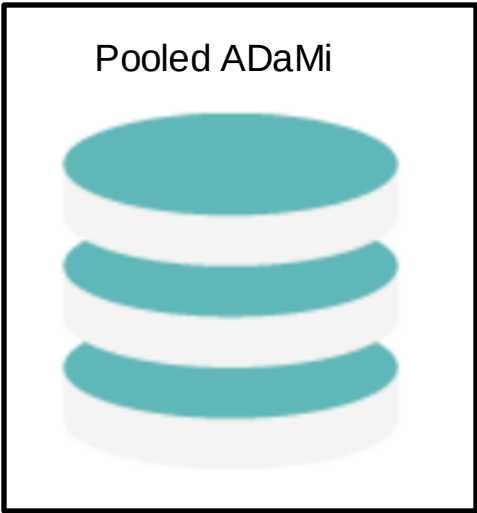
PARAM	PARAMCD	LBTESTCD	LABORATORY	LBTESTCD	LBSMPLID	LBRLOT
CSF: Amyloid Peptid Beta 40 (Elecsys Batch Neurotoolkit) (ng/mL)	CFAB40	AMYL B40	ELECSYS ROCHE DIAGNOSTICS NEUROTOOLKIT	CFAB40		
CSF: Amyloid Peptid Beta 40 (Elecsys at Blennow lab Batch at Screening) (ng/mL)	CFAB40E3	AMYL B40	ELECSYS ROCHE DIAGNOSTICS NEUROTOOLKIT	CFAB40E3		
CSF: Amyloid Peptid Beta 40 (Elecsys at Blennow lab Batch only patients with follow-up) (ng/mL)	CFAB40E4	AMYL B40	ELECSYS ROCHE DIAGNOSTICS NEUROTOOLKIT	CFAB40E4	ALIQUOT 1	
CSF: Amyloid Peptid Beta 40 (Lumipulse) (ng/mL)	CFAB40L	AMYL B40	LUMIPULSE FUJIREBIO	CFAB40L	ALIQUOT 2	
CSF: Amyloid Peptid Beta 40 (Elecsys Batch Neurotoolkit) (ng/mL) (Lot K05)	CFAB40L1	AMYL B40	ELECSYS ROCHE DIAGNOSTICS NEUROTOOLKIT	CFAB40L1		K05
CSF: Amyloid Peptid Beta 40 (Euroimmune) (ng/mL)	CFAB40M	AMYL B40	EUROIMMUNE	CFAB40M		
CSF: Amyloid Peptid Beta 40 (Eli Lilly Proprietary Assays) (ng/mL)	CFAB40P	AMYL B40	ELI LILLY PROPRIETARY ASSAYS	CFAB40P		
CSF: Amyloid Peptid Beta 40 (Elecsys Batch Neurotoolkit) (ng/mL) (Previous Lot)	CFAB40P1	AMYL B40	ELECSYS ROCHE DIAGNOSTICS NEUROTOOLKIT	CFAB40P1	PREVIOUS	
CSF: Amyloid Peptid Beta 40 (LCMS at PPD Batch) (ng/mL)	CFAB40PB	AMYL B40	INNOTEST FUJIREBIO	CFAB40PB		

Study Data versus Pooled Data

PET Visual Read

Additional values ??? for Variable PET Visual Read

PVISREAD	Subtotal	
	NO	
	YES	
	Ambiguous/indeterminate	
	Negative	
	Positive	



Study 1 - ADaM



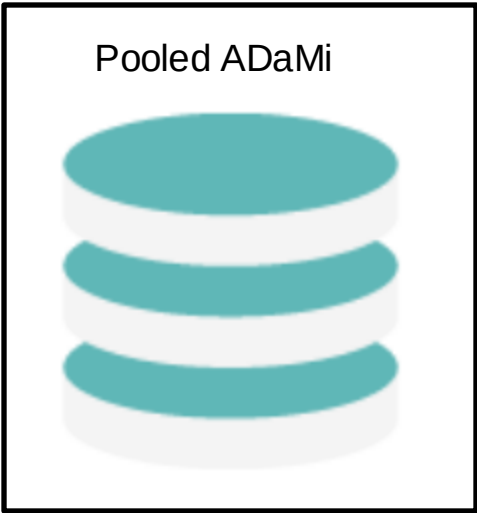
Study Data versus Pooled Data

PET Visual Read

Additional values ??? for Variable PET Visual Read

PVISREAD	Subtotal	
	NO	
	YES	
	Ambiguous/indeterminate	
	Negative	
	Positive	

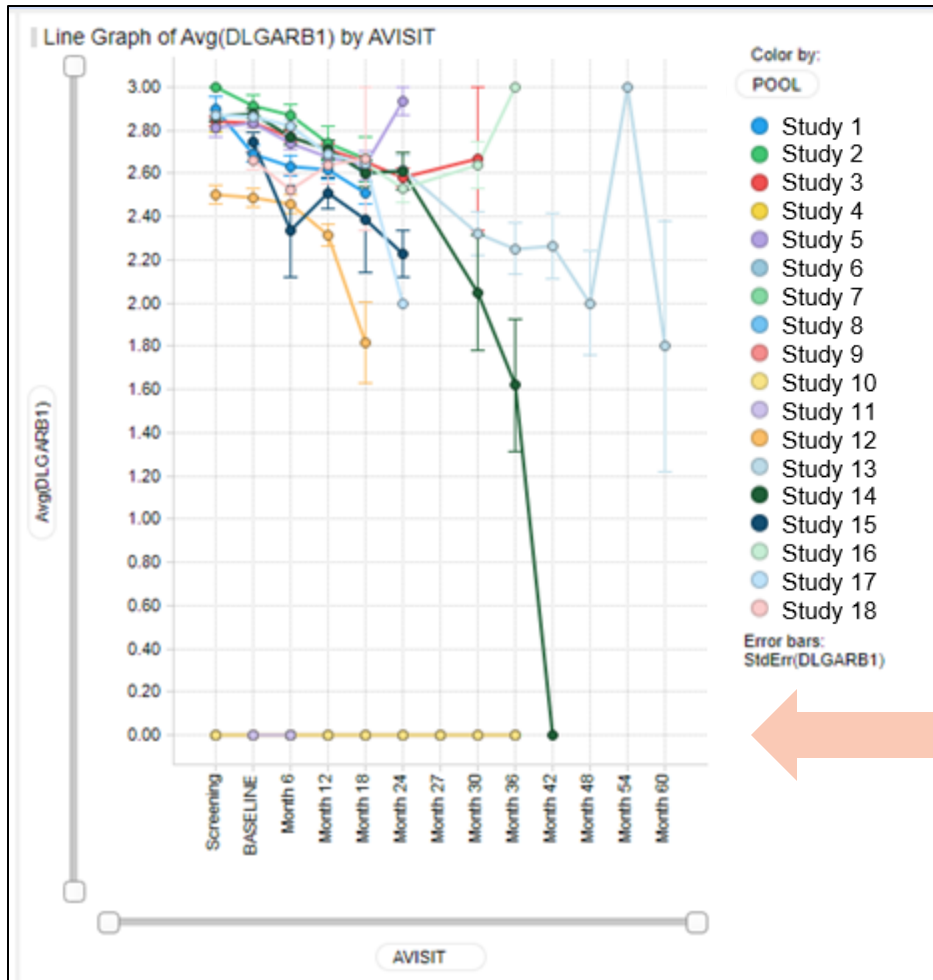
NO / YES are new values which were not included so far. Need to be harmonized into
-> Negative/ Positive



Study 1 - ADaM



Different [*TESTCD]



Yes No Don't Know

14. In the past 4 weeks, did {S} dispose of garbage or litter in an appropriate place or container at home?

If yes, which best describes how he/she usually performed:

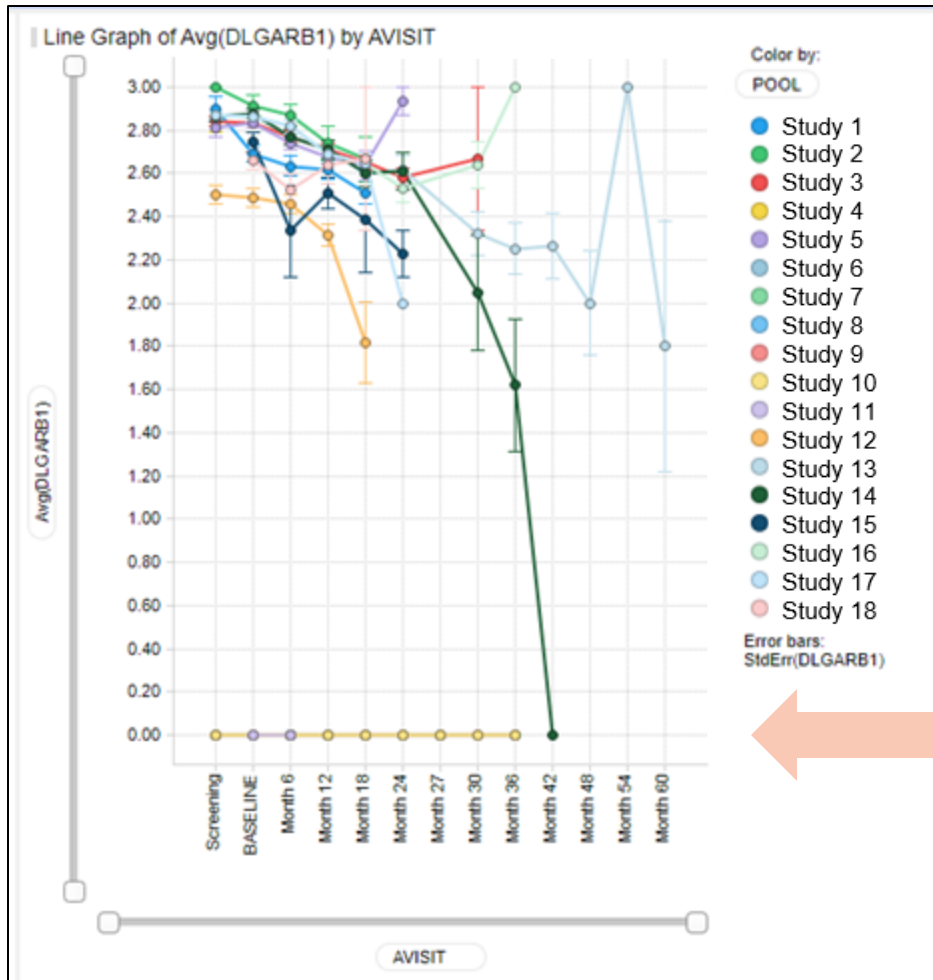
3 ☐ without supervision or help

2 ☐ with supervision

1 ☐ with physical help

Galasko D, Bennett D, Sano M, et al. An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. Alzheimer Dis Assoc Disord. 1997;11 Suppl 2:S33-9 ([PubMed abstract](#))

Different [*TESTCD]



Yes No Don't Know

14. In the past 4 weeks, did {S} dispose of garbage or litter in an appropriate place or container at home?

If yes, which best describes how he/she usually performed:

3 ☐ without supervision or help

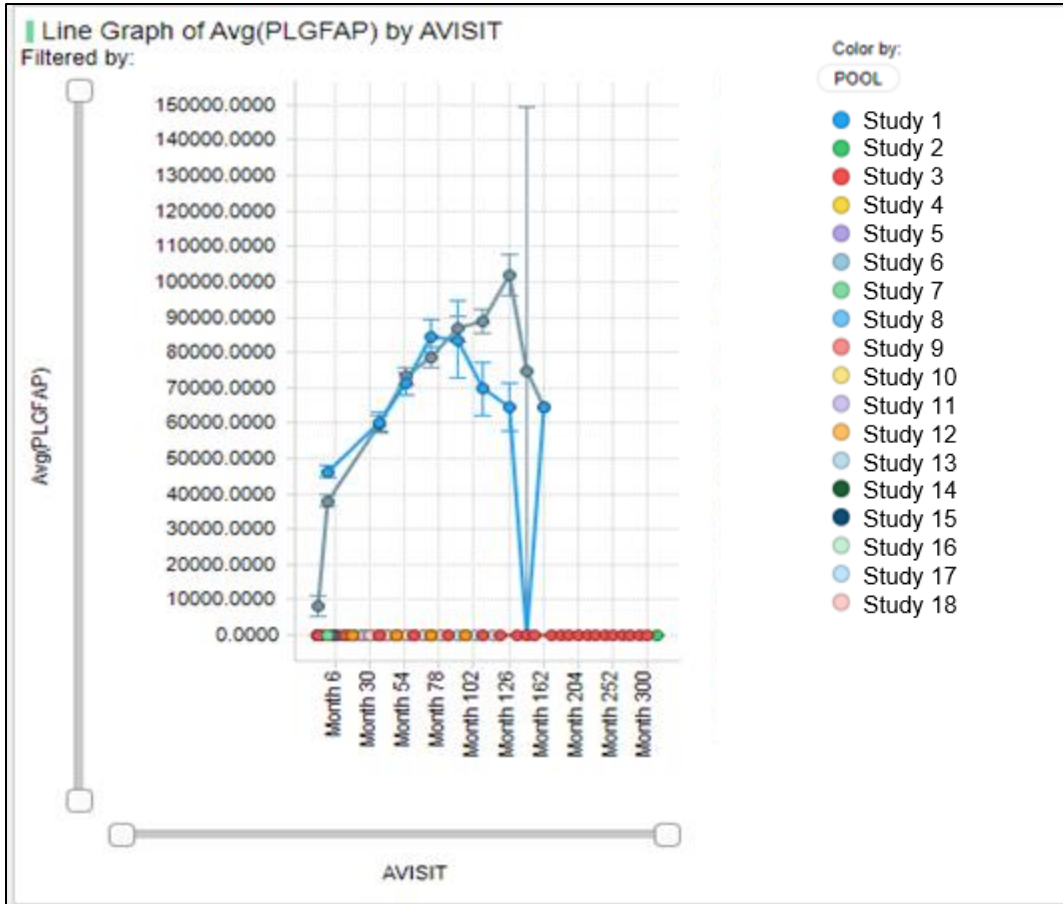
2 ☐ with supervision

1 ☐ with physical help

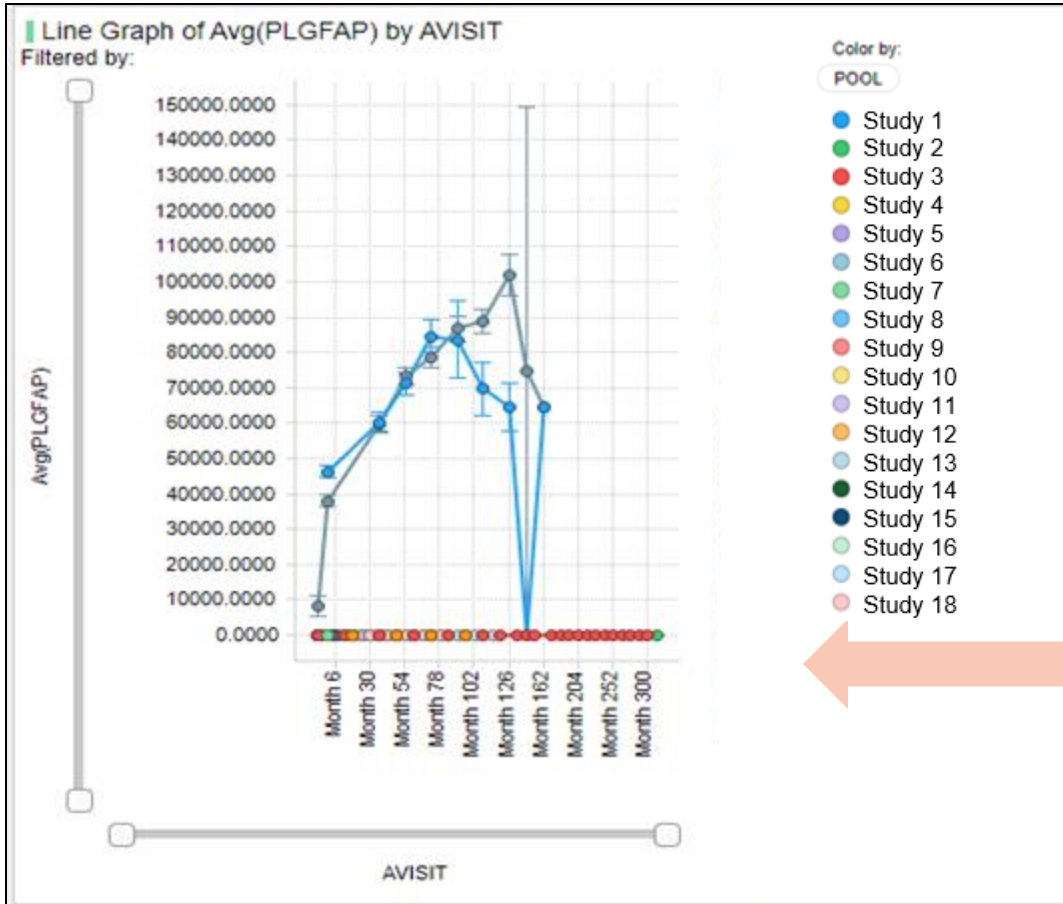
[QSTESTCD] mixed up:
«ADL0114A»

Galasko D, Bennett D, Sano M, et al. An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. Alzheimer Dis Assoc Disord. 1997;11 Suppl 2:S33-9 ([PubMed abstract](#))

Strange laboratory values in study data

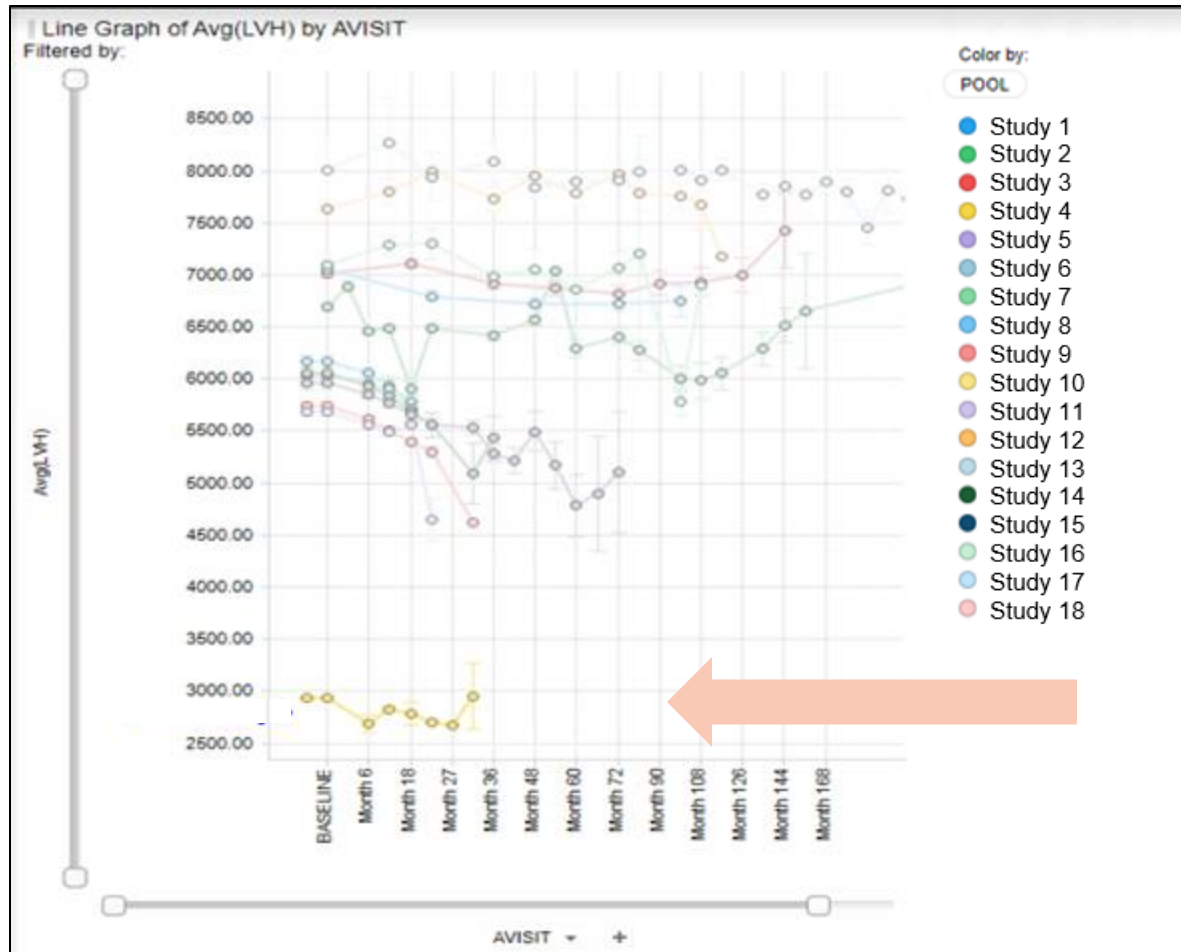


Strange laboratory values in study data



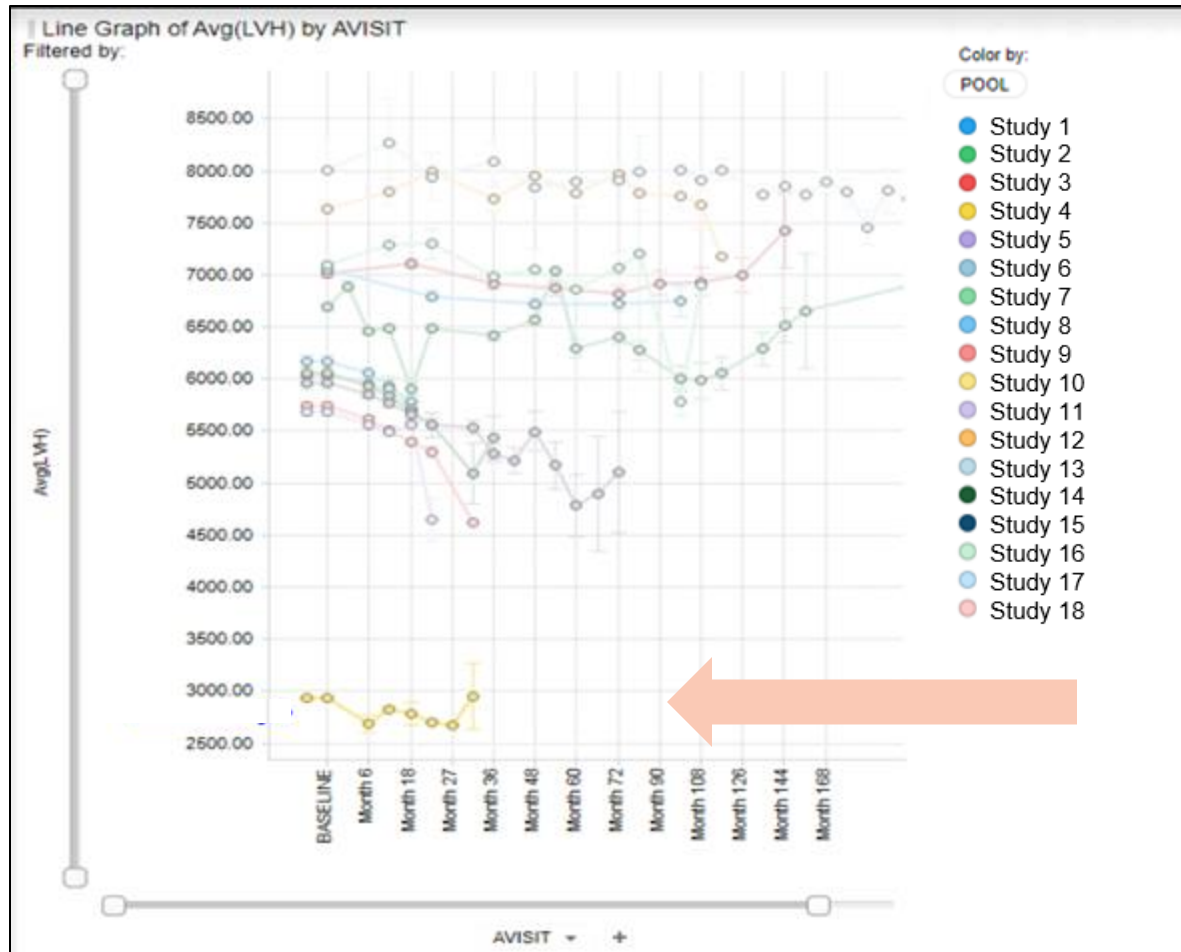
Laboratory unit
[LBSTRESU] issues with
original data from data
source
-> new data transfer

Data Issues: Volumetric MRI data



With comparing the data with other studies, the values look very low.

Data Issues: Volumetric MRI data



Hippocampus Volume was calculated as mean from left and right instead of sum within the source data.

Challenges Over Time

Updates of Standards within the company

Short Name

«COVA»

«FRLI»

«UGOT»

Long Name

«COVANCE»

«FRONTAGE LABORATORIES»,

«UNIVERSITY OF GOTHENBURG»

Upper Case

«ANN ROMNEY CENTER»

«COVANCE»

Lower Case

«Ann Romney Center»

«Covance»

Updates of SDTM standards

ApoE status

SDTM.SC (4 studies):

Stratification variable [SCORRES] = 0e4, 1e4, 2e4

SDTM.ZB (4 studies):

[ZB.ZBORRES] where [ZB.ZBTESTCD] = "ALLELE"

SDTM.PF:

[PF.PFSTRAIN] where [PF.PFTESTCD] = "ALLELE" and [PF.PFGENRI] = "APOE"

SDTM.GF (newest):

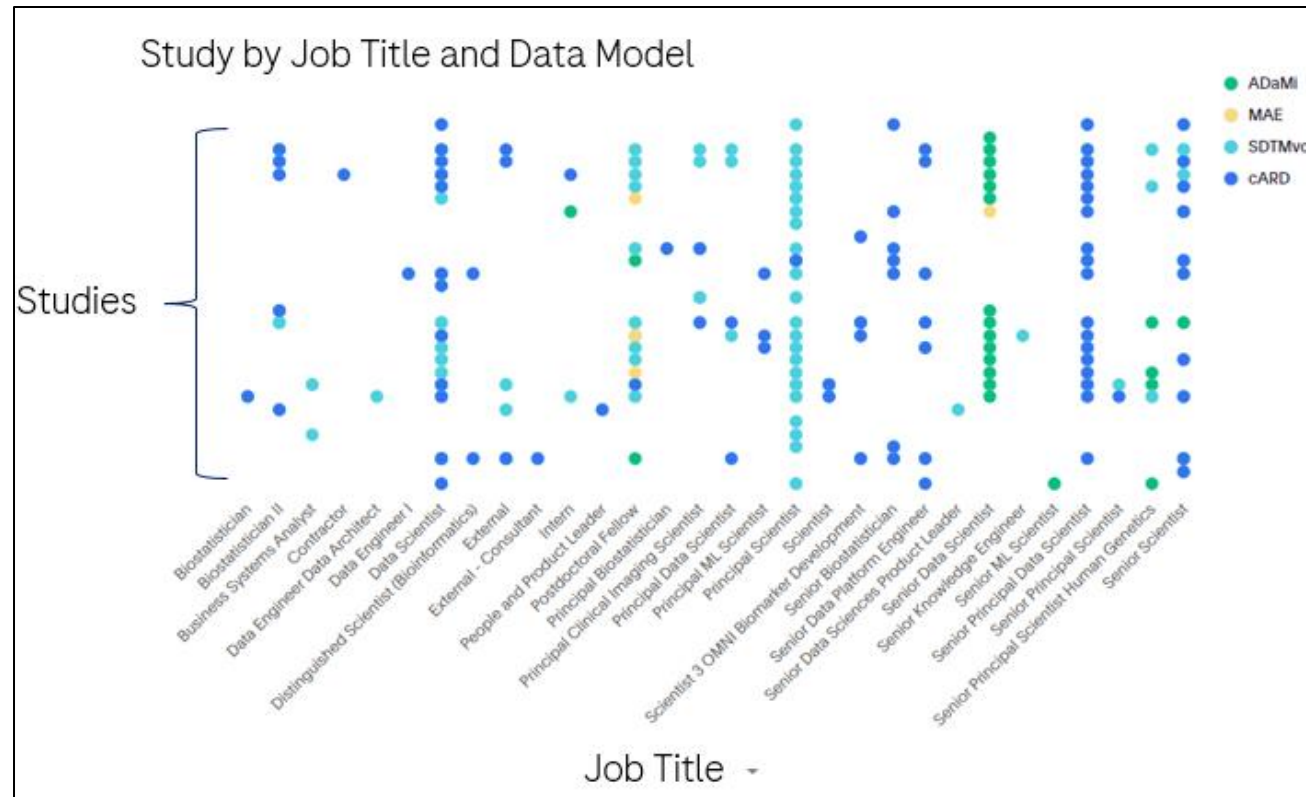
[GF.GFORRES] where [GF.GFTESTCD] = "ALLELE" and [GF.GFSYM] = "APOE"



Wins

Usage

>50 users across the organization



Feedback



It's so obvious to me....that would have been days and days of work to answer that question if we didn't have LeAD and I was able to sneak it in between meetings. The curation and harmonization work is so important.

Brad

"Trigger for exchange of statistical methods between various teams...."

Kirill

Yes, very helpful indeed! **Having to locate and pool the in-house studies done in AD ourselves would have been much more time consuming!** For example, we could use the pooled data to do the exploratory analyses needed to **build the early health economic model**

mart was helpful. I found the site ble summaries of the data (and n-outs") clear and helpful, and how to access and download the curate, and concise. This made a ocess a lot easier, and expedited ity".

Stephen

Feedback



It's so obvious to me....that would have been days and days of work to answer that question if we didn't have LeAD and I was able to sneak it in between meetings. The curation and harmonization work is so important.

Brad

“Trigger for exchange of statistical methods between various teams....”

Kirill

Yes, very helpful indeed! Having to locate and pool the in-house studies done in AD ourselves would have been much more time consuming! For example, we could use the pooled data to do the exploratory analyses needed to build the early health economic model

Sandro

“Yes, the data mart was helpful. I found the site intuitive, available summaries of the data (and possible "watch-outs") clear and helpful, and instructions on how to access and download the data simple, accurate, and concise. This made a complicated process a lot easier, and expedited the entire activity”.

Stephen

Feedback

It's so obvious to me....that would have been days and days of work to answer that **question if we didn't have LeAD and I was able to sneak it in between meetings.** The curation and harmonization work is so important.

model

Sandro

“Trigger for exchange of statistical methods between various teams...”

Kirill

“Yes, the data mart was helpful. I found the site intuitive, available summaries of the data (and possible "watch-outs") clear and helpful, and instructions on how to access and download the data simple, accurate, and concise. This made a complicated process a lot easier, and expedited the entire activity”.

Stephen

Feedback



It's so obvious to me....that would have been days and days of work to answer that question if we didn't have LeAD and I was able to sneak it in between meetings. The curation and harmonization work is so important.

Brad

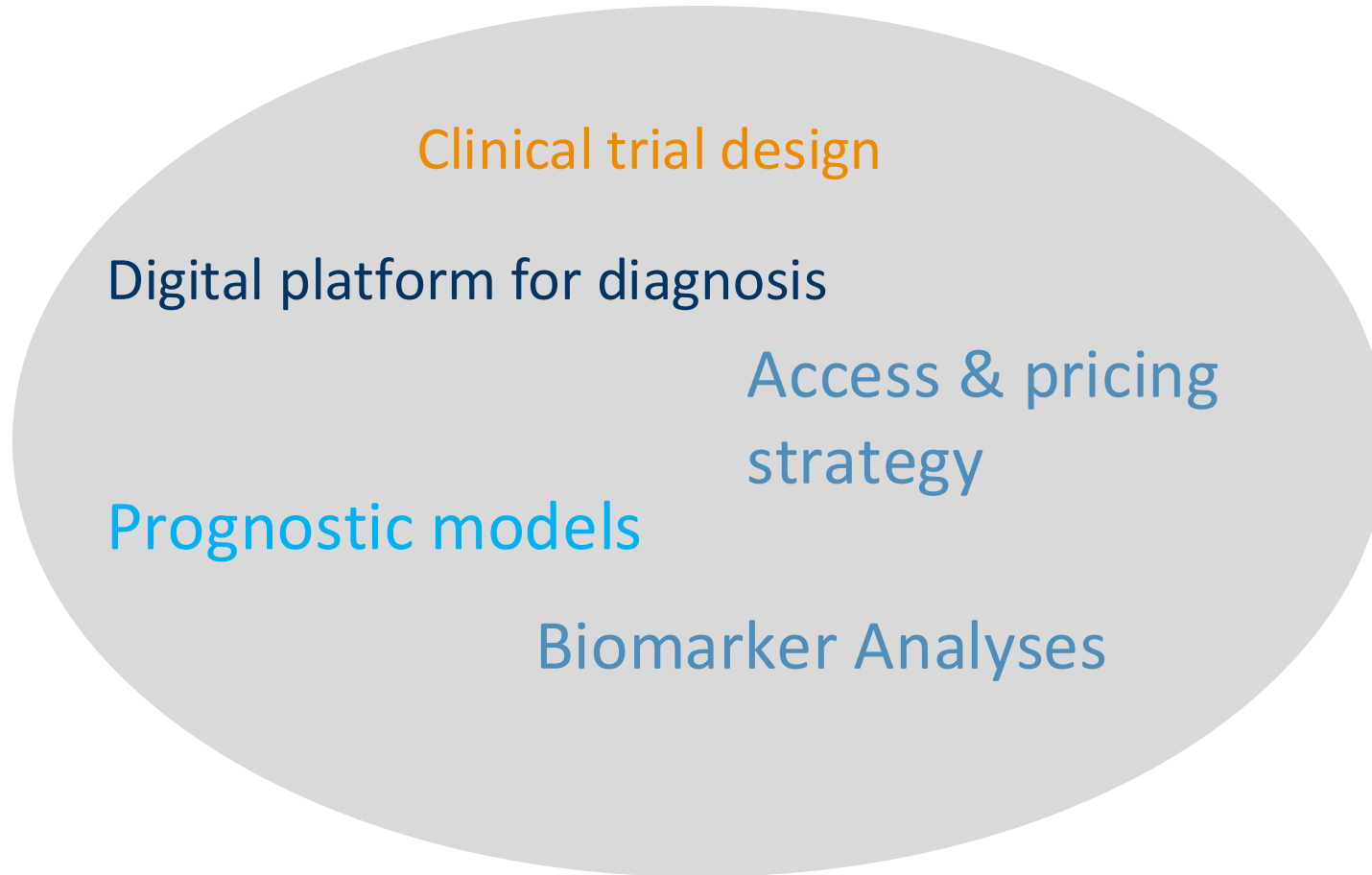
“Trigger for exchange of statistical methods between various teams...”

Kirill

Yes, very helpful indeed! Having to locate and pool the in-house studies done in AP...
consuming! For e...
exploratory anal
model

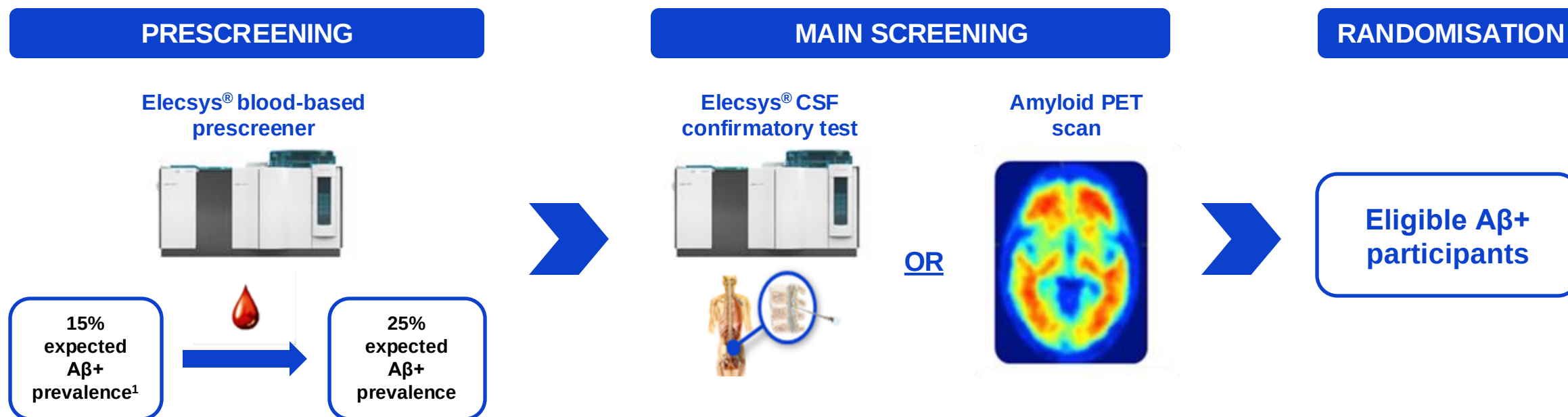
“Yes, the data mart was helpful. I found the site intuitive, available summaries of the data (and possible "watch-outs") clear and helpful, and instructions on how to access and download the data simple, accurate, and concise. This made a complicated process a lot easier, and expedited the entire activity”.

Large variety of use cases, impact and insights



Use of blood-based biomarker prescreening for SKYLINE secondary prevention study in Alzheimer's disease

One Roche approach to reduce clinical development costs and recruitment timelines



Optional Elecsys® blood-based prescreener (pTau181 and ApoE4 proteins)
to reduce screen failure rate during main screening and burden to participants and sites

Elecsys® blood-based prescreener is investigational and availability depends on local country / regulatory approval.
Elecsys® β-Amyloid (1-42) CSF and Elecsys® Phospho-Tau (181P) CSF are currently available in CE-mark accepting countries.
Aβ, beta-amyloid; ApoE4, apolipoprotein E ε4; CSF, cerebrospinal fluid; pTau, phosphorylated tau.
1. Sperling RA, et al. JAMA Neurol 2020;77:735–745.

Conclusions



Leveraging Exploratory Data and Analysis for Alzheimer's Disease

- FAIR Data
- Understanding of the data
- Stakeholder
- Secondary Use of Data

Any Questions?

Doing now what patients need next