Current Topics in Multi-Study Data Integration



Neda Jahanshad, PhD
Professor of Neurology, Biomedical Engineering, Information Sciences Institute
Director, LoBeS: Laboratory of Brain eScience
USC Mark and Mary Stevens Neuroimaging & Informatics Institute, Keck School of Medicine of USC
email: neda.jahanshad@ini.usc.edu

Laboratory of Brain eScience (LoBeS)

Supported by:

R01AG087513: *Causal and Event Based Modeling of Brain Alterations in ADRD*

R01MH134004: *Global studies into the Genetic Architecture of the Brain's White Matter Network through Harmonized and Coordinated Analyses in the ENIGMA-Consortium*

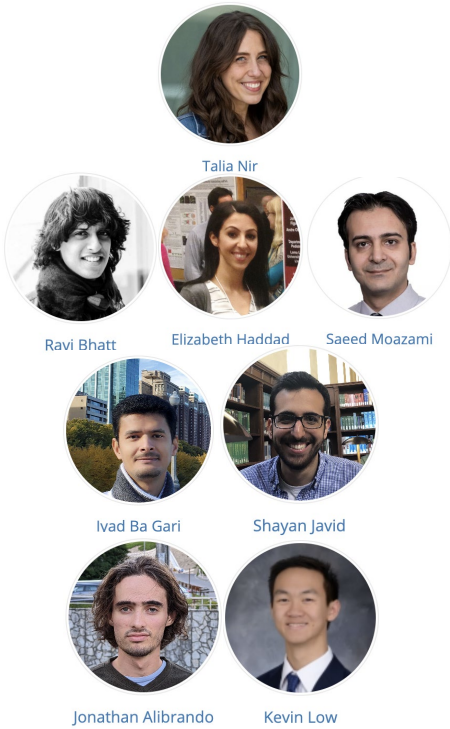
RF1NS136995: *Worldwide Tractometry Initiative to Investigate Brain Microstructure, Cognitive Impairment & Dementia in Parkinson's Disease*

R01HD117775: *Individual differences in brain morphology and connectivity to identify the mechanisms of comorbidities and treatment-refractory disease in females with endometriosis-associated pelvic pain*

R01NS128486: *Large-scale harmonization and integration of multi-modal ADNI data for the early detection of Alzheimer's disease and related dementias*



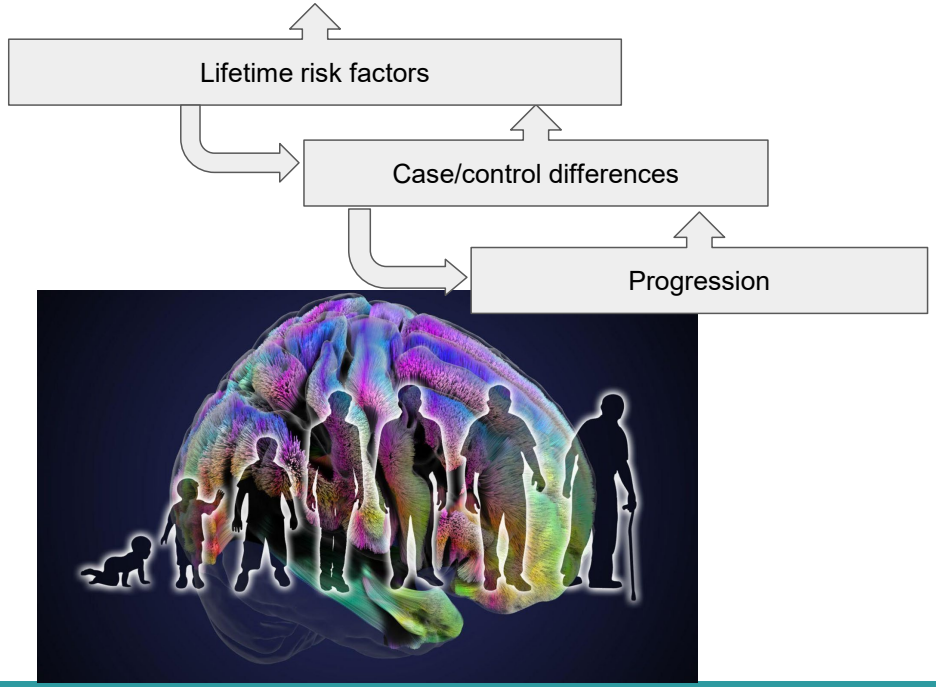
ALZHEIMER'S ASSOCIATION

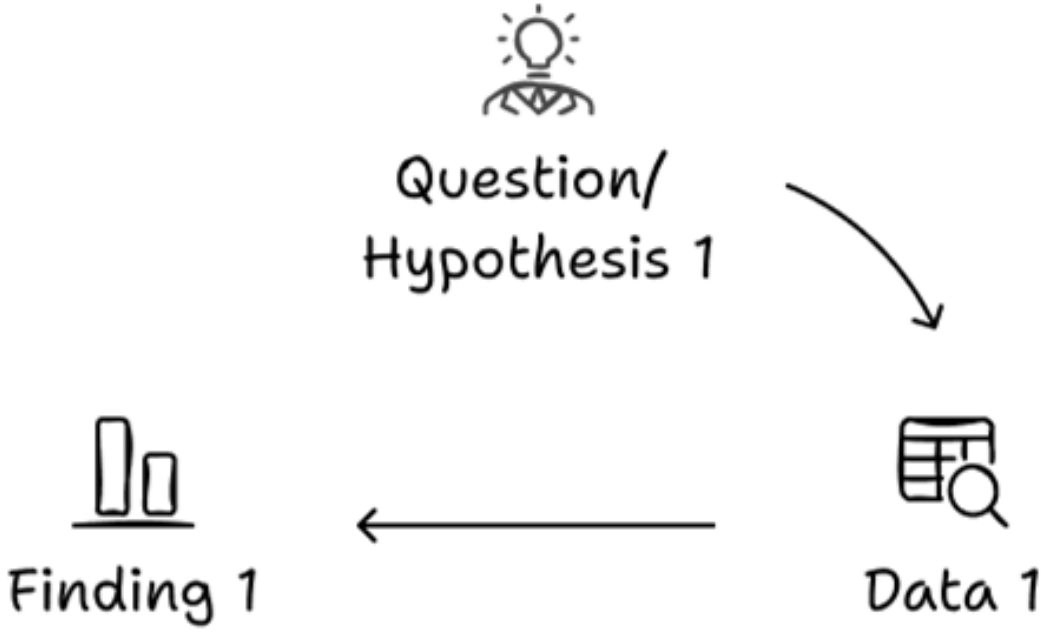


Ultimate Goal:

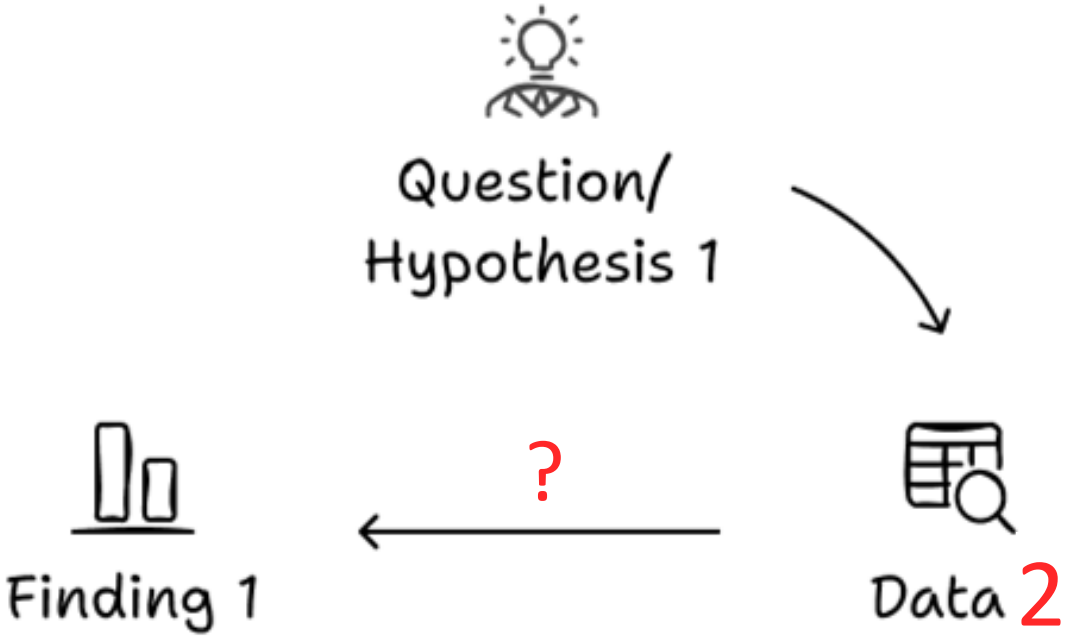
Optimizing individual brain health

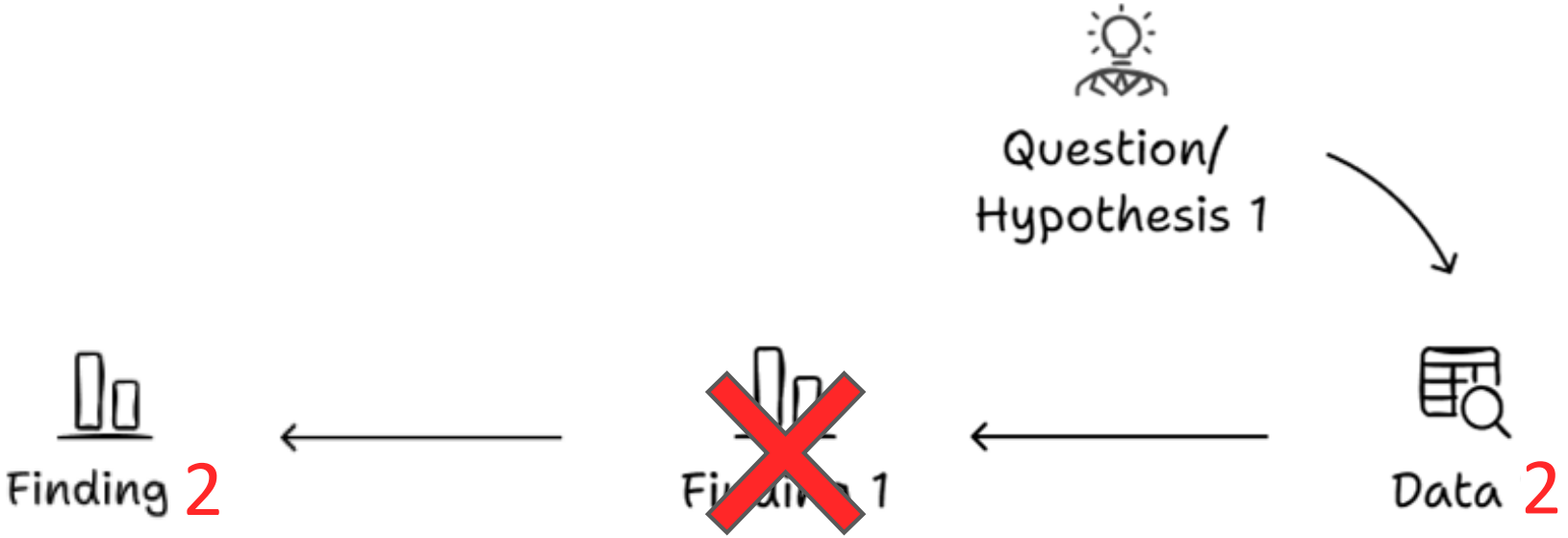
Understanding susceptibility to disease and identifying personal preventative strategies and treatments





Adapted from Jahanshad et al 2024


Question/
Hypothesis 1
?
Finding 1
Data 2


Question/
Hypothesis 1
Data 2
Finding 1
Finding 2

Headlines

Why Most Published Research Findings Are False

John P. A. Ioannidis

PLoS Medicine | www.plosmedicine.org 0696 August 2005 | Volume 2 | Issue 8 | e124

Power failure: why small sample size undermines the reliability of neuroscience

Katherine S. Button[1,2], John P. A. Ioannidis[3], Claire Mokrysz[1], Brian A. Nosek[4], Jonathan Flint[5], Emma S. J. Robinson[6] and Marcus R. Munafò[1]

NATURE REVIEWS | NEUROSCIENCE VOLUME 14 | MAY 2013 | 365

MOTHERBOARD
TECH BY VICE

Why Two Decades of Brain Research Could Be Seriously Flawed

A new paper casts doubt on 40,000 studies.

By **Kate Lunau**

Nature: Brain Imaging Studies Are Most Likely False

Small MRI studies inflate effect sizes, leaving the brain imaging research literature cluttered with false positives.

By **Peter Simons** - March 21, 2022

4013 16

Reproducibility in the neurosciences

The New York Times

Brain-Imaging Studies Hampered by Small Data Sets, Study Finds

Researchers have long used imaging technology to try to understand mental-health ailments. But with relatively few participants, such studies may

Give this article



Power failure: why small sample size undermines the reliability of neuroscience

Typical sample sizes < 50-100
Brain-wide associations may need 1000s of samples

Article | Published: 16 March 2022

Reproducible brain-wide association studies require thousands of individuals

Scott Marek, Brenden Tervo-Clemmens, Finnegan J. Calabro, David F. Montez, Benjamin P. Kay, Alexander S. Hatoum, Meghan Rose Donohue, William Foran, Ryland L. Miller, Timothy J. Hendrickson, Stephen M. Malone, Sridhar Kandala, Eric Feczko, Oscar Miranda-Dominguez, Alice M. Graham, Eric A. Earl, Anders J. Perrone, Michaela Cordova, Olivia Doyle, Lucille A. Moore, Gregory M. Conan, Johnny Uriarte, Kathy Snider, Benjamin J. Lynch, James C. Wilgenbusch, Thomas Pengo, Angela Tam, Jianzhong Chen, Dillan J. Newbold, Annie Zheng, Nicole A. Seider, Andrew N. Van, Athanasia Metoki, Roselyne J. Chauvin, Timothy O. Laumann, Deanna J. Greene, Steven E. Petersen, Hugh Garavan, Wesley K. Thompson, Thomas E. Nichols, B. T. Thomas Yeo, Deanna M. Barch, Beatriz Luna, Damien A. Fair & Nico U. F. Dosenbach — Show fewer authors

Nature **603**, 654–660 (2022) | Cite this article

15k Accesses | **1** Citations | **989** Altmetric | Metrics

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Reproducibility in the neurosciences

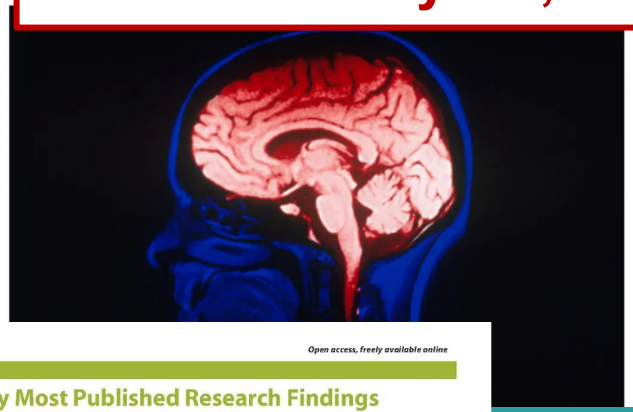
The New York Times

Brain-Imaging Studies Hampered by Small Data Sets, Study Finds

Researchers have long used imaging technology to try to understand mental-health ailments. But with relatively few

Power failure: why small sample size undermines the reliability of neuroscience

Need multi-site studies to increase participant count, power of statistical analyses, and establish greater confidence in findings.



Article | Published: 16 March 2022

Reproducible brain-wide association studies require thousands of individuals

Scott Marek, Brenden Tervo-Clemmens, Finnegan J. Calabro, David F. Montez, Benjamin P. Kay, Alexander S. Hatoum, Meghan Rose Donohue, William Foran, Ryland L. Miller, Timothy J. Hendrickson, Stephen M. Malone, Sridhar Kandala, Eric Feczko, Oscar Miranda-Dominguez, Alice M. Graham, Eric A. Earl, Anders J. Perrone, Michaela Cordova, Olivia Doyle, Lucille A. Moore, Gregory M. Conan, Johnny Uriarte, Kathy Snider, Benjamin J. Lynch, James C. Wilgenbusch, Thomas Pengo, Angela Tam, Jianzhong Chen, Dillan J. Newbold, Annie Zheng, Nicole A. Seider, Andrew N. Van, Athanasia Metoki, Roselyne J. Chauvin, Timothy O. Laumann, Deanna J. Greene, Steven E. Petersen, Hugh Garavan, Wesley K. Thompson, Thomas E. Nichols, B. T. Thomas Yeo, Deanna M. Barch, Beatriz Luna, Damien A. Fair & Nico U. F. Dosenbach — Show fewer authors

Nature **603**, 654–660 (2022) | Cite this article

15k Accesses | **1** Citations | **989** Altmetric | Metrics

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Multi Site Neuroimaging Studies

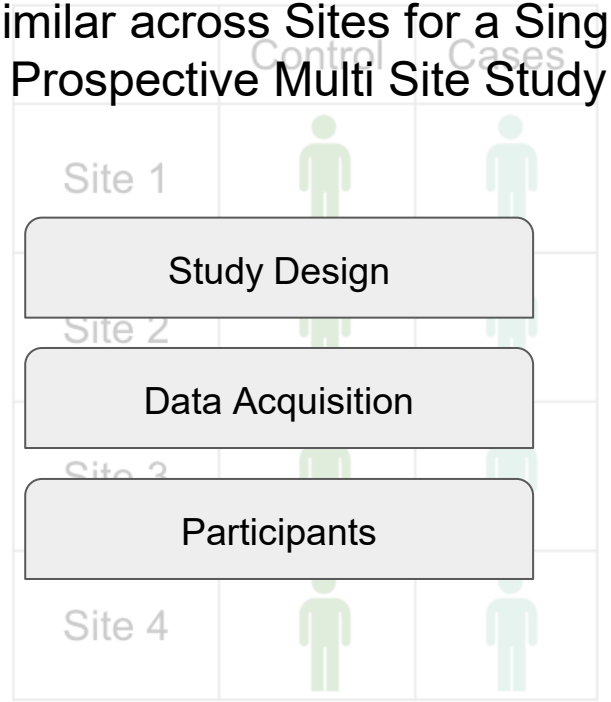
	Control	Cases
Study 1		

	Control	Cases
Site 1		
Site 2		
Site 3		
Site 4		

Multi Site Neuroimaging Studies

	Control	Cases
Study 1		

Similar across Sites for a Single Prospective Multi Site Study



Axes of data heterogeneity

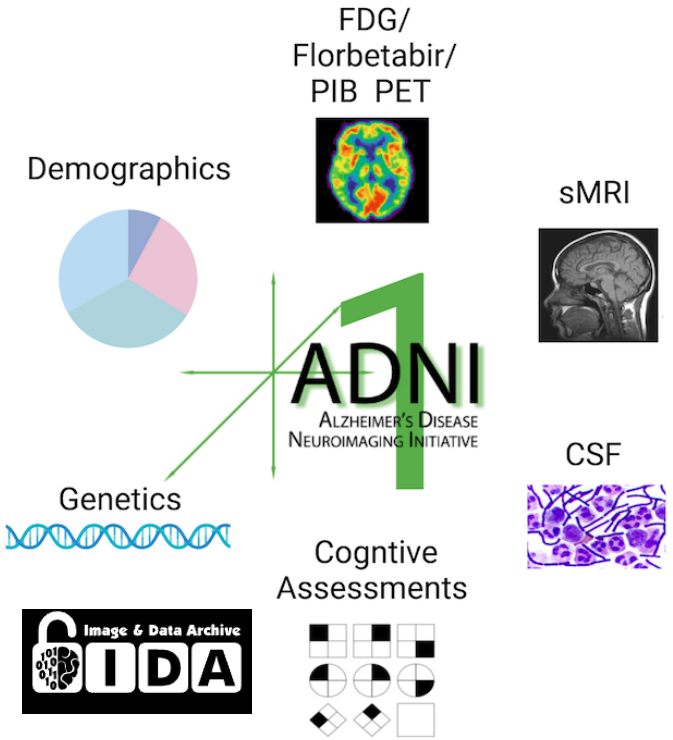
Usually, not

ADNI: Alzheimer's Disease Neuroimaging Initiative

Began in 2004 and is still ongoing

Phase 1 (2004 - 2010)

- **50+ sites around the US and Canada**
- Coordinated study designs and **harmonized MRI acquisition protocols**
  - Largest, most well powered, neuroimaging studies at the time
- Nearly 1000 subjects, multiple timepoints: Healthy controls, late-stage MCI, dementia
- Large scale initiative to identify demographic and biological risk factors for cognitive decline
- Publicly available for researchers around the world
  - Novel, harmonized storage and retrieval
  - ida.loni.usc.edu
- **Now completed Phase 4**



Multi Site Neuroimaging Studies -

Study Design Identical and Participants Similar across Sites
for a Single Prospective Multi Site Study

Of course, differences do exist

- Study Design
- **Data Acquisition**
- Participants

Neuroimaging:

Different scanners

Acquisition parameters

Challenge: Data Heterogeneity

Imaging acquisitions are different

Challenge: Data Heterogeneity



Challenge: Data Heterogeneity





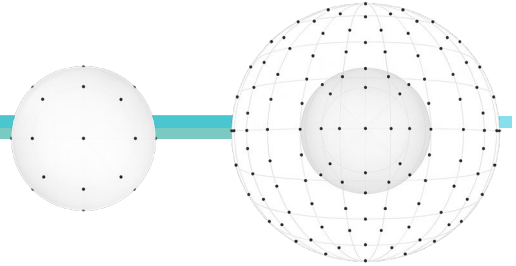
Challenge: Data Heterogeneity

Imaging acquisitions are different





Challenge: Data Heterogeneity



ADNI - MRI & diffusion MRI different in single study

	Protocol	Scanner		*Volumes*			Matrix	Time (min)
		Vendor	Protocol	b_0	DWI	Total		
ADNI3	GE36	GE	Basic Widebore 25x	4 b=0 s/mm^2	32 b=1000 s/mm^2	36	256x256	9:52
	GE54	GE	Basic 25x	6 b=0 s/mm^2	48 b=1000 s/mm^2	54	256x256	7:09
	P33	Philips	Basic Widebore	1 b=0 s/mm^2	32 b=1000 s/mm^2	33	128x128	7:32
	P36	Philips	Basic Widebore R3	1 b=0 s/mm^2 3 b=2 s/mm^2	32 b=1000 s/mm^2	36	128x128	6:54
	P54	Philips	Basic R5	1 b=0 s/mm^2 5 b=2 s/mm^2	48 b=1000 s/mm^2	54	128x128	8:05
	S31	Siemens	Basic VB17	1 b=0 s/mm^2	30 b=1000 s/mm^2	31	116x116	7:02
	S55	Siemens	Basic Skyra E11 & Prisma D13	7 b=0 s/mm^2	48 b=1000 s/mm^2	55	116x116	9:18
	S127	Siemens	Advanced Prisma VE11C	13 b=0 s/mm^2	6 b=500 s/mm^2 48 b=1000 s/mm^2 60 b=2000 s/mm^2	127	116x116	7:25*

Growing neuroimaging resources



HABS-HD

hsc Institute for Translational Research

biobank uk

Improving the health of future generations

OASIS
OPEN ACCESS SERIES OF IMAGING STUDIES

Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD)

aibl

camCAN
Cambridge Centre for Ageing and Neuroscience

CONNECTOME
COORDINATION FACILITY

Alzheimer's Disease Connectome Project

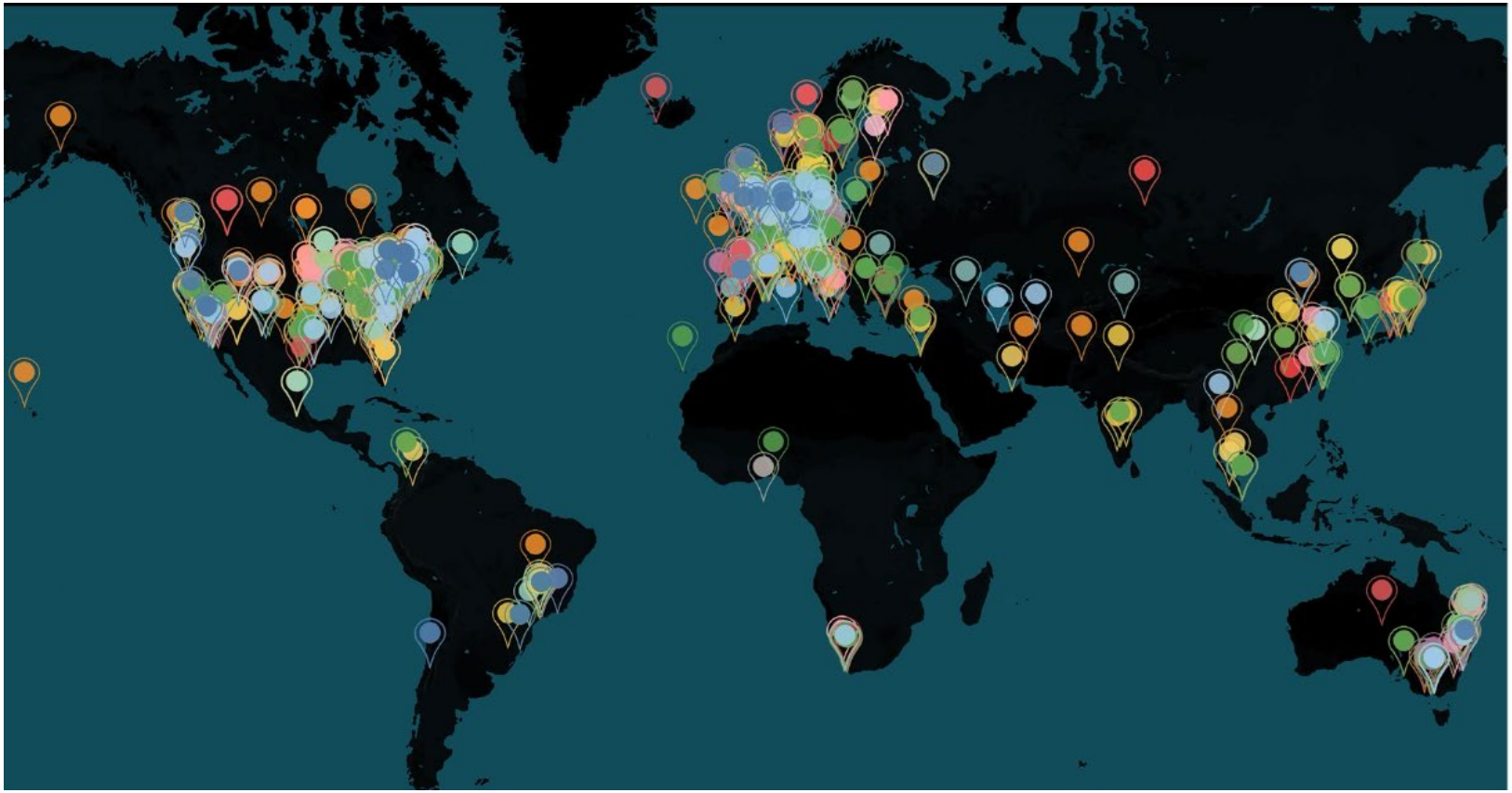
NACC

THE NIA ALZHEIMER'S DISEASE RESEARCH CENTERS PROGRAM
National Alzheimer's Coordinating Center

PREVENT-AD
Research Group

HARVARD AGING BRAIN
Massachusetts General Hospital – Harvard Medical School – Brigham and Women's Hospital

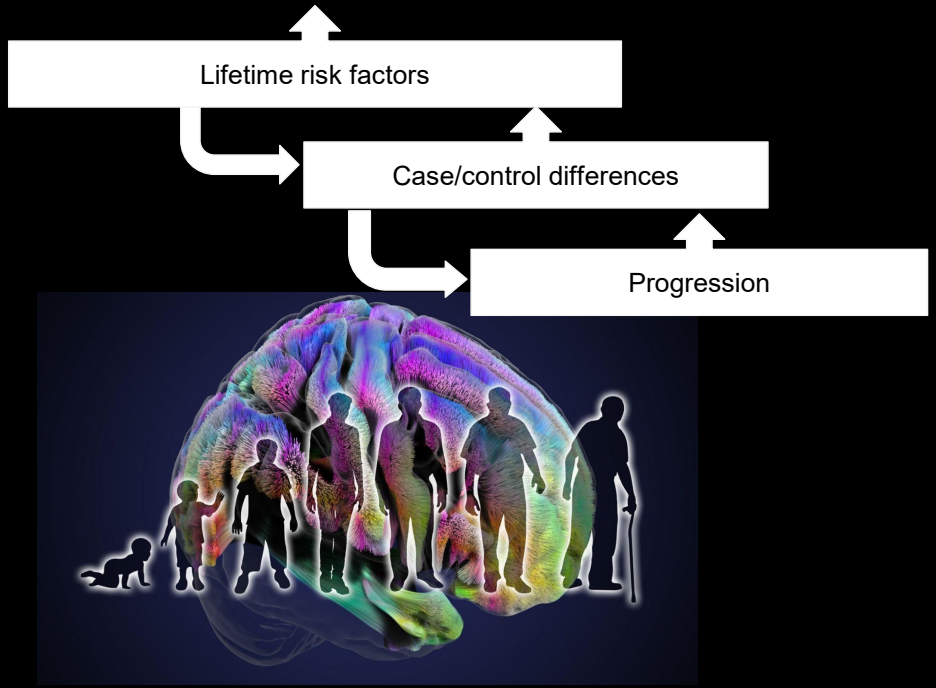
Decades of Neuroimaging Studies around the World



Ultimate Goal:

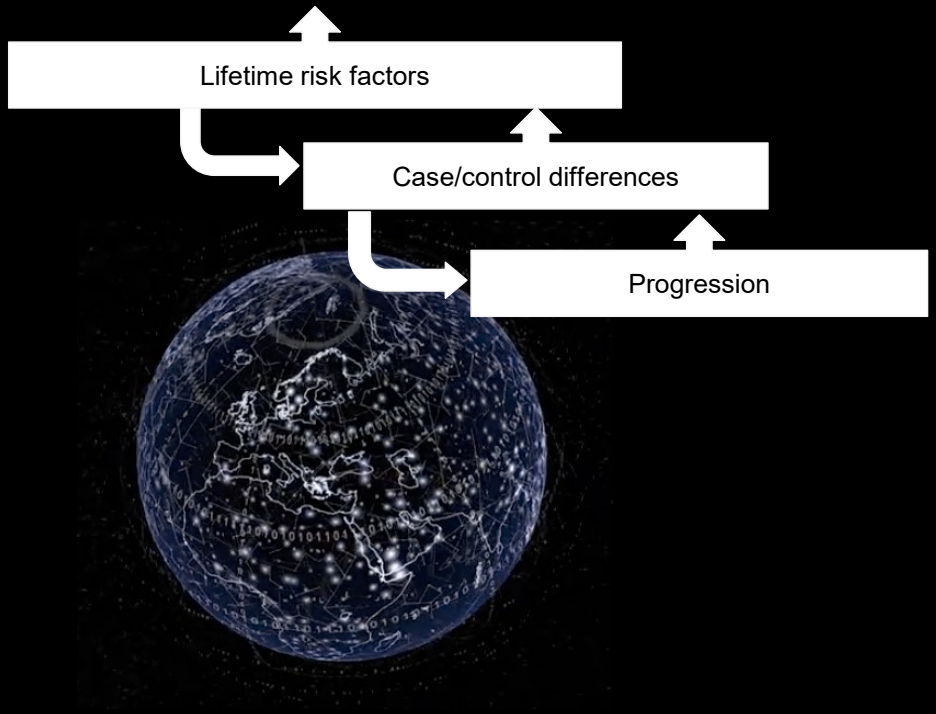
Optimizing individual brain health

Understanding susceptibility to disease and identifying personal preventative strategies and treatments



… In populations around the world

Understanding susceptibility to disease and identifying personal preventative strategies and treatments



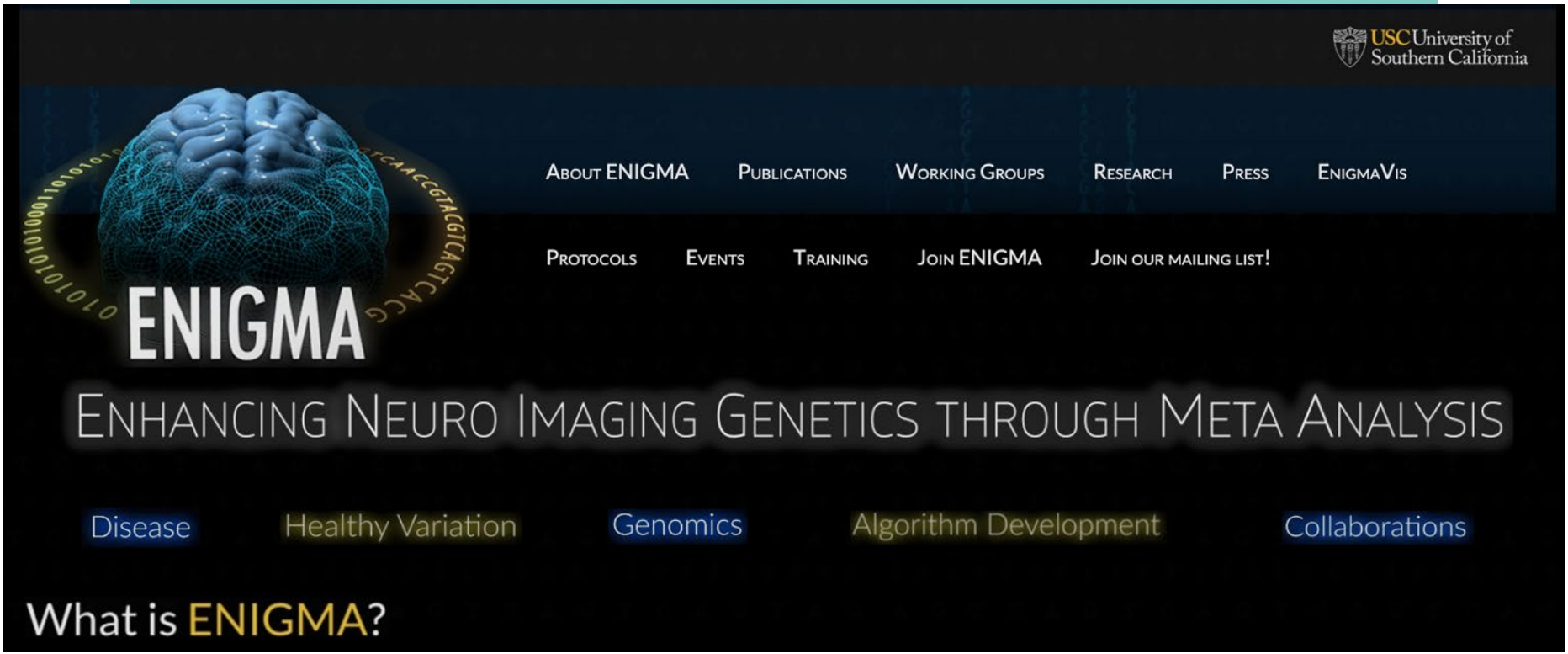
Multi ***Study*** Neuroimaging Studies

	Control	Cases
Study 1		
Study 2		
Study 3		
Study 4		

	Control	Cases
Study 1 Site 1		
Study 1 Site 2		
Study 2 Site 1		
Study 3 Site 1		
Study 3 Site 2		
Study 4 Site 1		

All independent / different

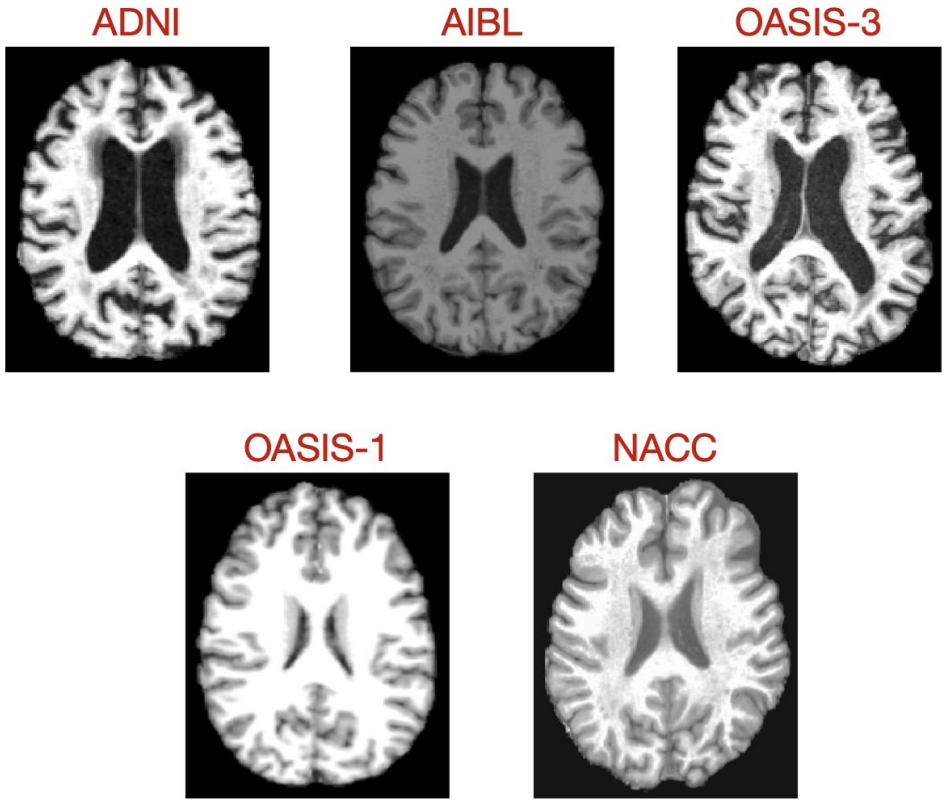
- Study Design
- Data Acquisition
- Participants



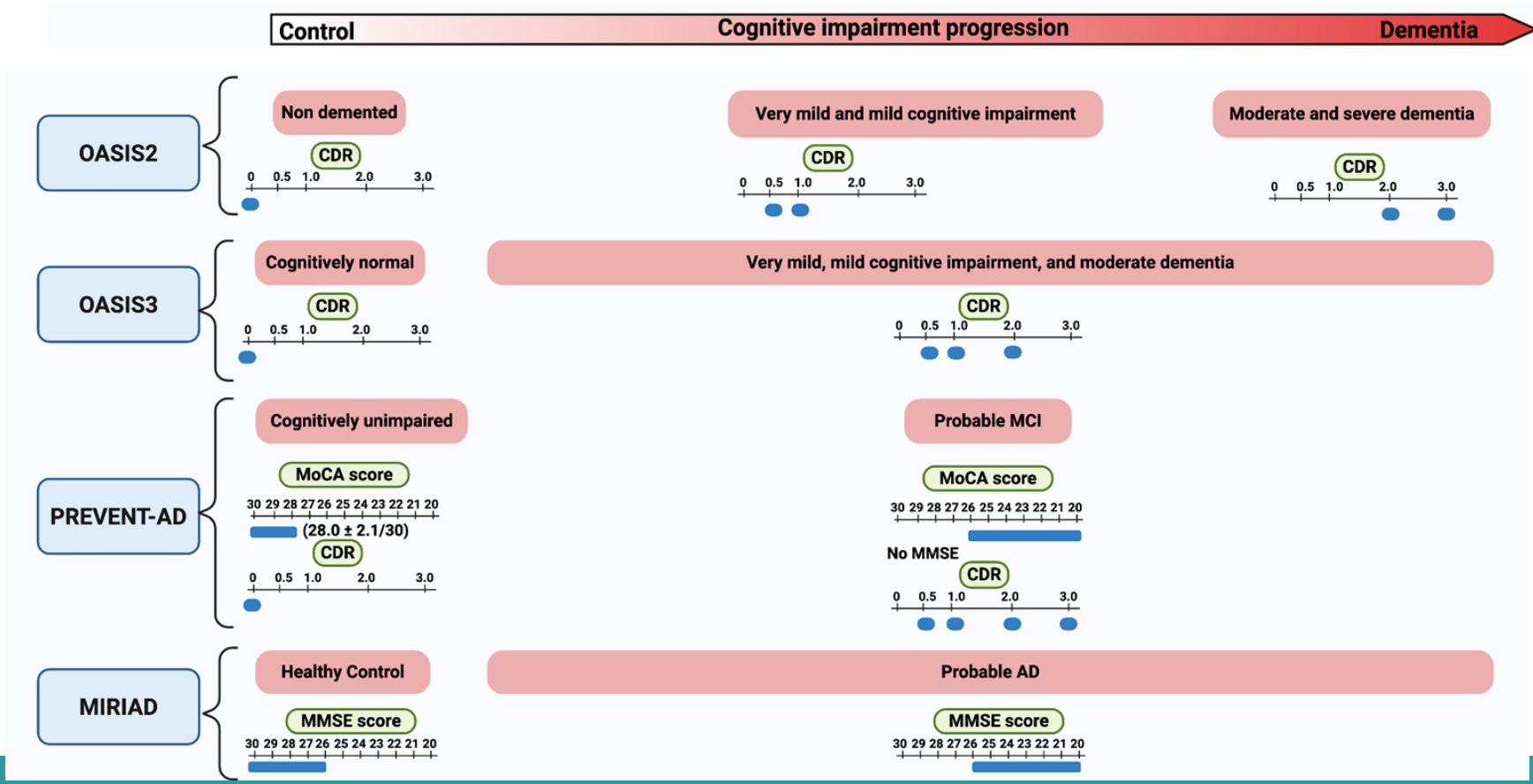
Distributed scientific collaboration. Site related heterogeneity weighted.
Effects are weighted by the variability in the site level data / or its sample size

Thompson, Jahanshad et al Trans Psych 2020 for a review

Imaging variability



Diagnostic variability



ADNI exclusion criteria

EXCLUSION CRITERIA			
Item	**NL**	**MCI**	**AD**
Significant neurologic disease	Any significant neurologic disease, such as Parkinson's disease, multi-infarct dementia, Huntington's disease, normal pressure hydrocephalus, brain tumor, progressive supranuclear palsy, seizure disorder, subdural hematoma, multiple sclerosis, or history of significant head trauma followed by persistent neurologic defaults or known structural brain abnormalities.	Any significant neurologic disease other than suspected incipient Alzheimer's disease, such as Parkinson's disease, multi-infarct dementia, Huntington's disease, normal pressure hydrocephalus, brain tumor, progressive supranuclear palsy, seizure disorder, subdural hematoma, multiple sclerosis, or history of significant head trauma followed by persistent neurologic defaults or known structural brain abnormalities.	Any significant neurologic disease other than Alzheimer's disease including Parkinson's disease, multi-infarct dementia, Huntington's disease, normal pressure hydrocephalus, brain tumor, progressive supranuclear palsy, seizure disorder, subdural hematoma, multiple sclerosis, or history of significant head trauma followed by persistent neurologic defaults or known structural brain abnormalities.
Neuroimaging	Screening/baseline MRI scans with evidence of infection, infarction, or other focal lesions. Subjects with multiple lacunes or lacunes in a critical memory structure are excluded.	Screening/baseline MRI scans with evidence of infection, infarction, or other focal lesions. Subjects with multiple lacunes or lacunes in a critical memory structure are excluded.	Screening/baseline MRI scans with evidence of infection, infarction, or other focal lesions. Subjects with multiple lacunes or lacunes in a critical memory structure are excluded.
MRI exclusions	Presence of pacemakers, aneurysm clips, artificial heart valves, ear implants, metal fragments or foreign objects in the eyes, skin or body.	Presence of pacemakers, aneurysm clips, artificial heart valves, ear implants, metal fragments or foreign objects in the eyes, skin or body.	Presence of pacemakers, aneurysm clips, artificial heart valves, ear implants, metal fragments or foreign objects in the eyes, skin or body.

EXCLUSION CRITERIA			
Item	**NL**	**MCI**	**AD**
Psychiatric disorders/psychotic features	Major depression, bipolar disorder as described in DSM-IV within the past 1 year. History of schizophrenia (DSM IV criteria).	Major depression, bipolar disorder as described in DSM-IV within the past 1 year. Psychotic features, agitation or behavioral problems within the last 3 months which could lead to difficulty complying with the protocol. History of schizophrenia (DSM IV criteria).	Major depression, bipolar disorder as described in DSM-IV within the past 1 year. Psychotic features, agitation or behavioral problems within the last 3 months which could lead to difficulty complying with the protocol. History of schizophrenia (DSM IV criteria).
Alcohol abuse	History of alcohol or substance abuse or dependence within the past 2 years (DSM IV criteria).	History of alcohol or substance abuse or dependence within the past 2 years (DSM IV criteria).	History of alcohol or substance abuse or dependence within the past 2 years (DSM IV criteria).
Significant medical illness	Any significant systemic illness or unstable medical condition which could lead to difficulty complying with the protocol.	Any significant systemic illness or unstable medical condition which could lead to difficulty complying with the protocol.	Any significant systemic illness or unstable medical condition which could lead to difficulty complying with the protocol.
Clinically significant laboratory abnormalities	Clinically significant abnormalities in B12, RPR, or TFTs that might interfere with the study.	Clinically significant abnormalities in B12, RPR, or TFTs that might interfere with the study.	Clinically significant abnormalities in B12, RPR, or TFTs that might interfere with the study.
Residence	Residence in skilled nursing facility.	Residence in skilled nursing facility.	Residence in skilled nursing facility.

ADNI *excludes*

EXCLUSION CRITERIA

Item	NL	MCI	AD
Significant neurologic disease	Any significant neurologic disease, such as Parkinson's disease, multi-infarct dementia, Huntington's disease, normal pressure hydrocephalus, brain tumor, progressive supranuclear palsy, seizure disorder, subdural hematoma, multiple sclerosis, or history of significant head trauma followed by persistent neurologic defaults or known structural brain abnormalities.	Any significant neurologic disease other than suspected incipient disease, multi-infarct dementia, Huntington's disease, normal pressure hydrocephalus, brain tumor, progressive supranuclear palsy, seizure disorder, subdural hematoma, multiple sclerosis, or history of significant head trauma followed by persistent neurologic defaults or known structural brain abnormalities.	Any significant neurologic disease other than Alzheimer's disease dementia, Huntington's disease, normal pressure hydrocephalus, brain tumor, progressive supranuclear palsy, seizure disorder, subdural hematoma, multiple sclerosis, or history of significant head trauma followed by persistent neurologic defaults or known structural brain abnormalities.
Neuroimaging	Screening/baseline MRI scans with evidence of infection, infarction, or other focal lesions. Subjects with multiple lacunes or lacunes in a critical memory structure are excluded.	Screening/baseline MRI scans with evidence of infection, infarction, or other focal lesions. Subjects with multiple lacunes or lacunes in a critical memory structure are excluded.	Screening/baseline MRI scans with evidence of infection, infarction, or other focal lesions. Subjects with multiple lacunes or lacunes in a critical memory structure are excluded.
MRI exclusions	Presence of pacemakers, aneurysm clips, artificial heart valves, ear implants, metal fragments or foreign objects in the eyes, skin or body.	Presence of pacemakers, aneurysm clips, artificial heart valves, ear implants, metal fragments or foreign objects in the eyes, skin or body.	Presence of pacemakers, aneurysm clips, artificial heart valves, ear implants, metal fragments or foreign objects in the eyes, skin or body.

Progressive supranuclear palsy

Parkinson's disease

History of significant head trauma

Seizure disorder

EXCLUSION CRITERIA

Item	NL	MCI	AD
Psychiatric disorders/psychotic features	Major depression, bipolar disorder as described in DSM-IV within the past 1 year. History of schizophrenia (DSM IV criteria).	Major depression, bipolar disorder as described in DSM-IV within the past 1 year. psychotic features, agitation or behavioral problems within the last 3 months which could lead to difficulty History of schizophrenia (DSM IV criteria).	Major depression, bipolar disorder as described in DSM-IV within the past 1 year. psychotic features, agitation or behavioral problems within the last 3 months which could lead to difficulty complying with the protocol. History of schizophrenia (DSM IV criteria).
Alcohol abuse	History of alcohol or substance abuse or	History of alcohol or substance abuse or	History of alcohol or substance abuse or the IV criteria).
Significant medical illness	Any significant systemic illness or unstable medical condition which could lead to difficulty complying with the protocol.	Any significant systemic illness or unstable medical condition which could lead to difficulty complying with the protocol.	Any significant systemic illness or unstable medical condition which could lead to difficulty complying with the protocol.
Clinically significant laboratory abnormalities	Clinically significant abnormalities in B12, RBC, or TFTs that might interfere with the study.	Clinically significant abnormalities in B12, RBC, or TFTs that might interfere with the study.	Clinically significant abnormalities in B12, RBC, or TFTs that might interfere with the study.
Residence	Residence in skilled nursing facility.	Residence in skilled nursing facility.	Residence in skilled nursing facility.

Major depression

Schizophrenia

History of alcohol or substance abuse

Vitamin B12 deficiency

OASIS 3 *includes…*

CN Co-occurs with	AD Dementia	AD dem w/depresss, not cor
Active Affective disorder	0.5 in memory only	0.5 in memory only
Active Alcoholism	Active Alcoholism	Active Alcoholism
Active B-12 Deficiency	Active B-12 Deficiency	Active B-12 Deficiency
Active Bereavement	Active Bereavement	Active CVD
Active CVD	Active CVD	Active Hypothyroidism
Active Hypothyroidism	Active Hydrocephalus	Active Major Head Trauma
Active Major Head Trauma	Active Hypothyroidism	Active Med-induced Cog Dys
Active Med-induced Cog Dys	Active Med-induced Cog Dys	Active Mood disorder
Active Mood disorder	Active Mood disorder	Active Other neurol/med diagnoses
Active Other neurol/med diagnoses	Active Other neurol/med diagnoses	Active PD idiopathic
Active PD drug-induced	Active PD drug-induced	AD dem distrubed social, with
Active PD idiopathic	Active PD idiopathic	AD dem Language dysf after
Active Seizure Disorder	Active Seizure Disorder	AD dem Language dysf prior
Frontotemporal demt. prim	AD dem distrubed social, after	AD dem Language dysf with
Oth potentially transmissible dis	AD dem distrubed social, with	AD dem visuospatial, with
other neurol/med diagnoses	AD dem Language dysf after	AD dem w/CVD contribut
ProAph w/o dement	AD dem Language dysf prior	AD dem w/CVD not contrib
Remote Alcoholism	AD dem Language dysf with	AD dem w/oth (list B) contribut
Remote Amnestic syndrome	AD dem visuospatial, after	AD dem w/oth (list B) not contrib
Remote B-12 Deficiency	AD dem visuospatial, prior	AD dem w/PDI after AD dem contribut
Remote Bereavement	AD dem visuospatial, with	AD dem w/PDI after AD dem not contrib
Remote CVD	AD dem w/CVD contribut	AD Dementia
Remote Global Cerebral Hypop	AD dem w/CVD not contrib	Frontotemporal demt. secn
Remote Hydrocephalus	AD dem w/depresss, contribut	Incipient demt PTP
Remote Hypothyroidism	AD dem w/depresss, not contribut	Remote Alcoholism
Remote Major Head Trauma	AD dem w/oth (list B) contribut	Remote B-12 Deficiency
Remote Med-induced Cog Dys	AD dem w/oth (list B) not contrib	Remote CVD
Remote Mood disorder	AD dem w/PDI after AD dem contribu	Remote Hypothyroidism
Remote Other neurol/med diagnoses	AD dem w/PDI after AD dem not con	Remote Major Head Trauma
Remote PD drug-induced	DLBD, secondary	Remote Med-induced Cog Dys
Remote PD idiopathic	other mental retarAD demion	Remote Mood disorder
Remote Seizure Disorder	Park Idiopathic	Remote Other neurol/med diagnoses
	Rapid Course (not CJD)	Remote Seizure Disorder
	Remote Alcoholism	uncertain dementia
	Remote B-12 Deficiency	
	Remote Bereavement	

- AD dem cannot be primary
- AD dem distrubed social, after
- AD dem distrubed social, prior
- AD dem distrubed social, with
- AD dem Language dysf after
- AD dem Language dysf prior
- AD dem Language dysf with
- AD dem visuospatial, after
- AD dem visuospatial, prior
- AD dem visuospatial, with
- AD dem w/CVD contribut
- AD dem w/CVD not contrib
- AD dem w/depresss, contribut
- AD dem w/depresss, not contribut
- AD dem w/Frontal lobe/demt at onset
- AD dem w/oth (list B) contribut
- AD dem w/oth (list B) not contrib
- AD dem w/oth unusual feat/subs demt
- AD dem w/oth unusual features
- AD dem w/oth unusual features/demt on
- AD dem w/PDI after AD dem contribut
- AD dem w/PDI after AD dem not contrib
- AD dem/FLD prior to AD dem
- AD Dementia

OASIS 3 *includes…* (in cognitively normal and impaired)

CN Co-occurs with	AD Dementia	AD dem w/depresss, not cor
Active Affective disorder	0.5 in memory only	0.5 in memory only
Active Alcoholism	Active Alcoholism	Active Alcoholism
Active B-12 Deficiency	Active B-12 Deficiency	Active B-12 Deficiency
Active Bereavement	Active Bereavement	Active CVD
Active CVD	Active CVD	Active Hypothyroidism
Active Hypothyroidism	Active Hydrocephalus	Active Major Head Trauma
Active Major Head Trauma		
Active Med-induced Cog Dys		
Active Mood disorder	Active Mood disorder	Active Other neurol/med diagnoses
Active Other neurol/med diagnoses	Active Other neurol/med diagnoses	Active PD idiopathic
	Active PD drug-induced	AD dem distrubed social, with
	Active PD idiopathic	AD dem Language dysf after
	Active Seizure Disorder	AD dem Language dysf prior
Frontotemporal demt. prin	AD dem distrubed social, after	AD dem Language dysf with
Oth potentially transmissible	AD dem distrubed social, prior	AD dem visuospatial, with
other neurol/med diagnoses	AD dem Language dysf after	AD dem w/CVD contribut
ProAph w/o dement	AD dem Language dysf prior	AD dem w/CVD not contrib
Remote Alcoholism	AD dem Language dysf with	AD dem w/oth (list B) contribut
Remote Amnestic Syn	AD dem visuospatial, after	AD dem w/oth (list B) not contrib
Remote B-12 Deficiency	AD dem visuospatial, prior	AD dem w/PDI after AD dem contribut
Remote Bereavement	AD dem visuospatial, with	AD dem w/PDI after AD dem not contrib
Remote CVD	AD dem w/CVD contribut	AD Dementia
Remote Global Cerebral Hypop	AD dem w/CVD not contrib	Frontotemporal demt. secn
Remote Hydrocephalus	AD dem w/depresss, contribut	Incipient demt PTP
Remote Hypothyroidism	AD dem w/depresss, not contribut	Remote Alcoholism
Remote Major Head Trauma	AD dem w/oth (list B) contribut	Remote B-12 Deficiency
Remote Med-induced Cog Dys	AD dem w/oth (list B) not contrib	Remote CVD
Remote Mood disorder	AD dem w/PDI after AD dem contribut	Remote Hypothyroidism
Remote Other neurol/med diagnoses	AD dem w/PDI after AD dem not con	Remote Major Head Trauma
Remote PD drug-induced	DLBD, secondary	Remote Med-induced Cog Dys
Remote PD idiopathic	other mental retarAD demion	Remote Mood disorder
Remote Seizure Disorder	Park Idiopathic	Remote Other neurol/med diagnoses
	Rapid Course (not CJD)	Remote Seizure Disorder
	Remote Alcoholism	uncertain dementia
	Remote B-12 Deficiency	
	Remote Bereavement	

AD dem cannot be primary
AD dem distrubed social, after
AD dem distrubed social, prior
AD dem distrubed social, with
AD dem Language dysf after
AD dem Language dysf prior
AD dem Language dysf with
AD dem visuospatial, after
AD dem visuospatial, prior
AD dem visuospatial, with
AD dem w/CVD contribut
AD dem w/CVD not contrib
AD dem w/depresss, contribut
AD dem w/depresss, not contribut
AD dem w/oth (list B) contribut
AD dem w/oth (list B) not contrib
AD dem w/oth unusual feat/subs demt
AD dem w/oth unusual features
AD dem w/oth unusual features/demt on
AD dem w/PDI after AD dem contribut
AD dem w/PDI after AD dem not contrib
AD dem/FLD prion
AD Dementia

Progressive supranuclear palsy

Parkinson's disease

History of significant head trauma

Seizure disorder

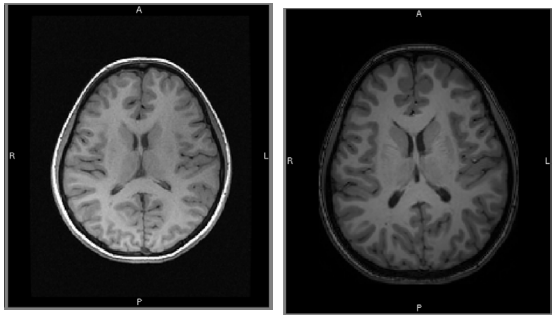
Major depression

History of alcohol or substance abuse

Vitamin B12 deficiency

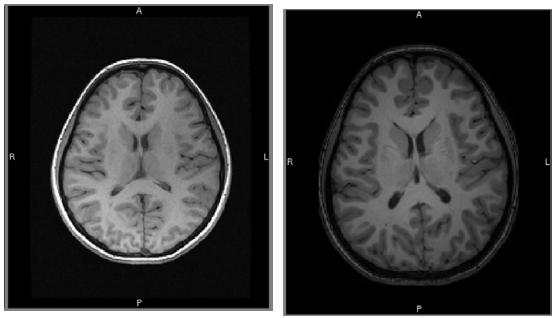
Challenge: Image Data heterogeneity

Imaging acquisitions are different



Challenge: Image Data heterogeneity

Imaging acquisitions are different



that's ok!

A biomarker may have limited clinical utility NOW if it is only evident when measured with instruments that are not widely available.

Do not want a biomarker to rely on one particular instrument/scanner for global impact.

Benefits of Heterogeneity: Reproducibility, Reliability, and **Generalizability**

With large diverse, heterogeneous datasets, we can readily assess generalizability.

Will the findings hold in a different population?
With a different method: acquisition / analytical ?

If not, why? Biological? Methodological?

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Whitaker K. 2017. Publishing a reproducible paper. figshare doi: 10.6084/m9.figshare.5440621.v2
https://the-turing-way.netlify.app/reproducible-research/

Approach: Analytic Standardization

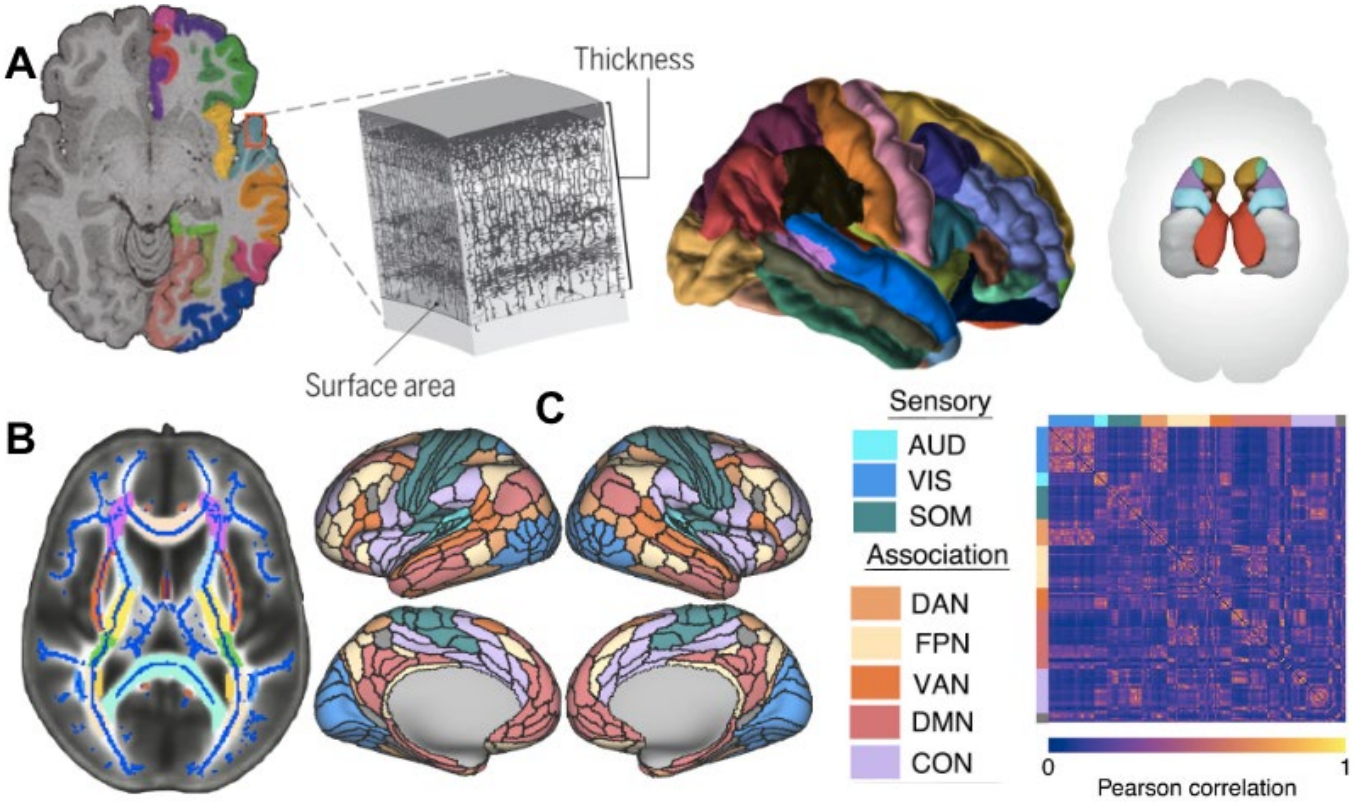
Image Processing / Feature Extraction Workflows in ENIGMA

Our general approach to **pooling diverse data** has largely revolved around minimizing variability through analytic standardization

Detailed and **documented analytic plans** are key

- **Imaging metrics**
- **Quality assurance**
- **Statistical workflow**
- **Test / retest / evaluate**

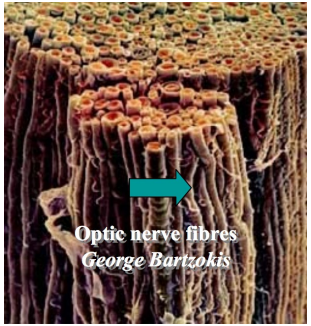
Standardized Imaging Pipelines

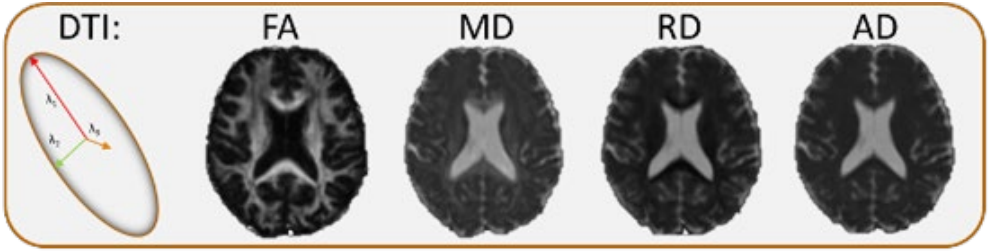


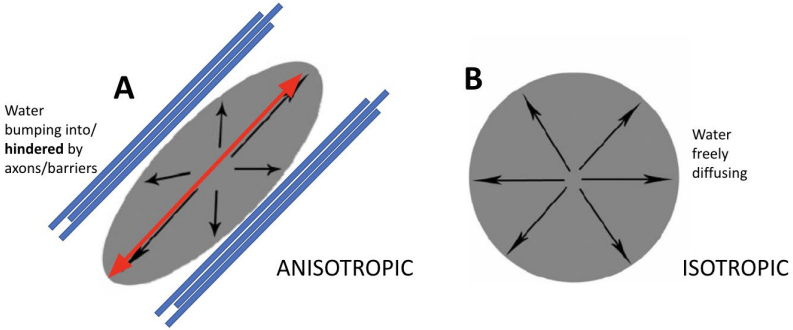
Diffusion MRI based microstructure

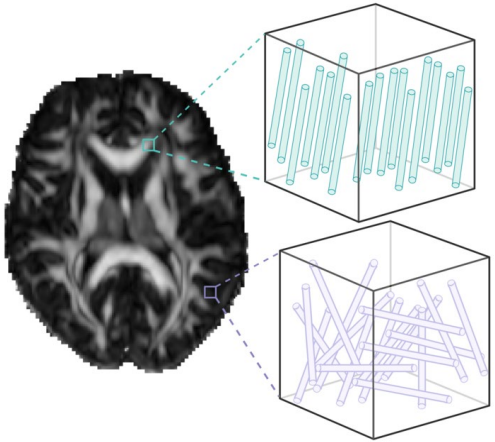
4D - Diffusion MRI - DTI: basic model









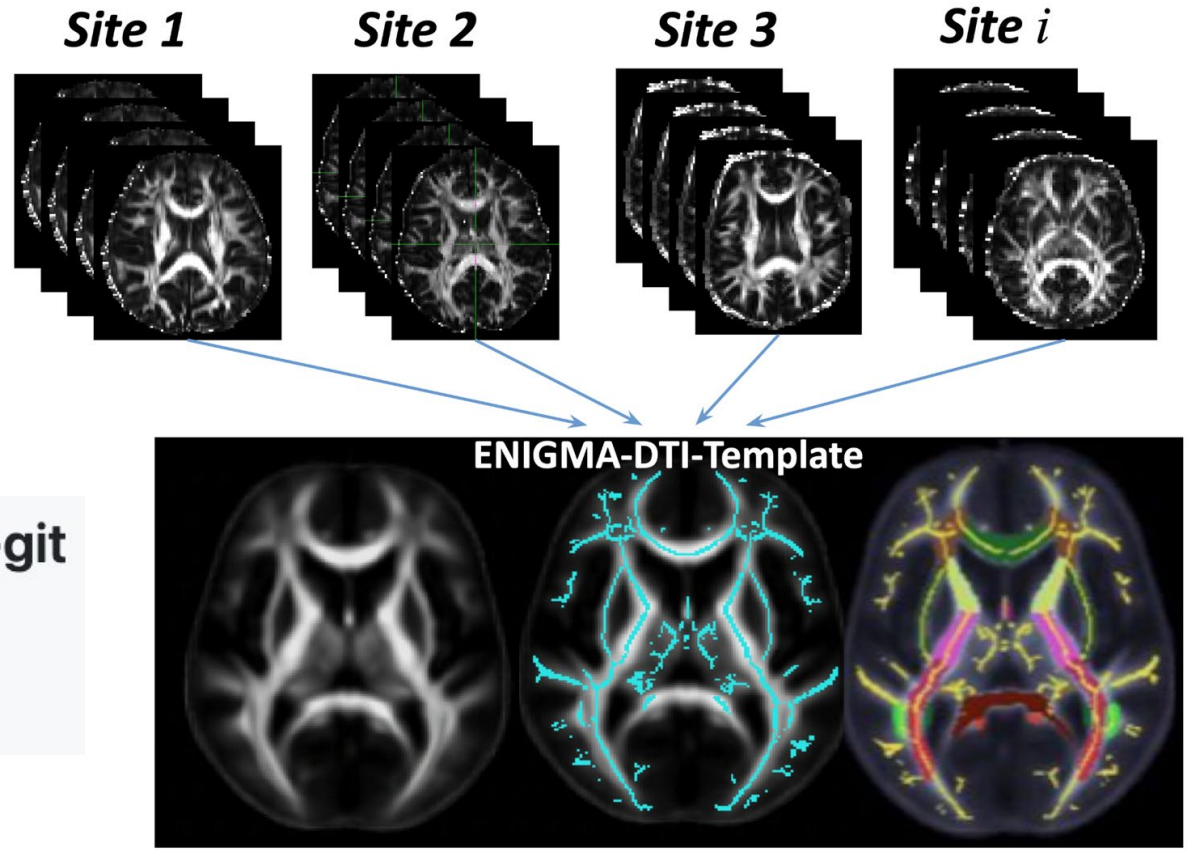


Estimate of water motion and directionality

Obtain relative estimates of cohesion and dispersion

Standardization of dMRI Imaging Methods

White matter microstructure

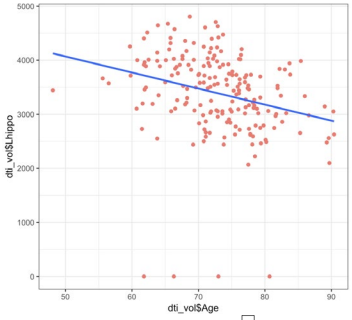


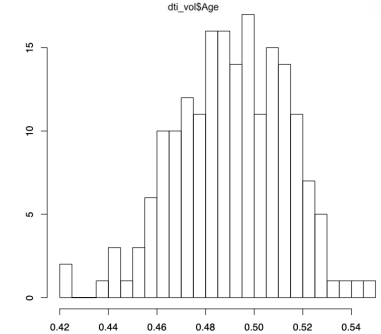
ENIGMA-git

Jahanshad et al 2013;
Builds on FSL's TBSS (Smith et al 2006)

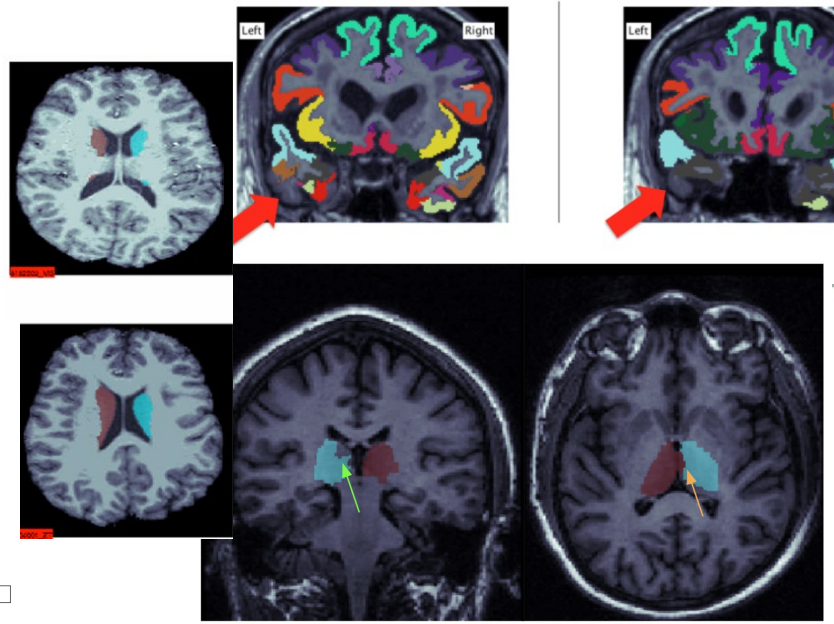
Analytical Standardization: Quality Assurance

The documentation of quality assurance procedures is a critical step

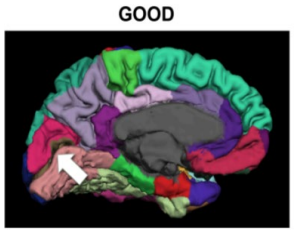




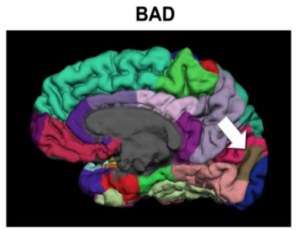
Internal QC: Moderate
Temporal Pole Underestimation



Pericalcarine Overestimations (<5% cases)



- Segmentation confined to calcarine sulcus

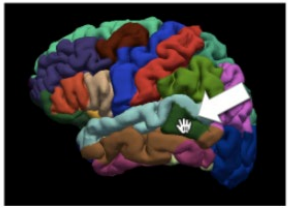


- Segmentation overestimates pericalcarine region

Bad examples BanksSTS
(List affected regions in QC sheet)



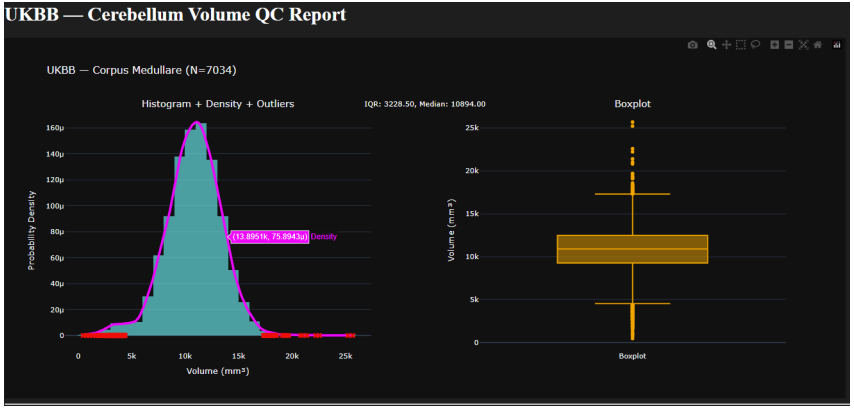
Fail BanksSTS + STG

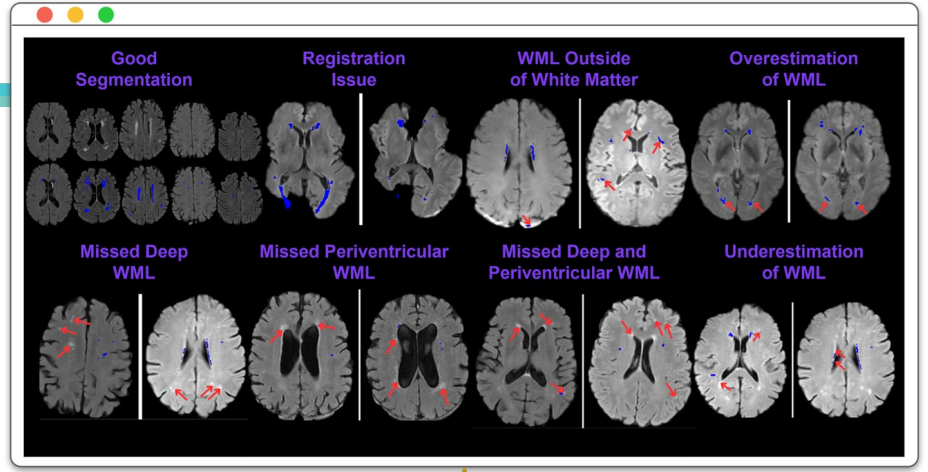


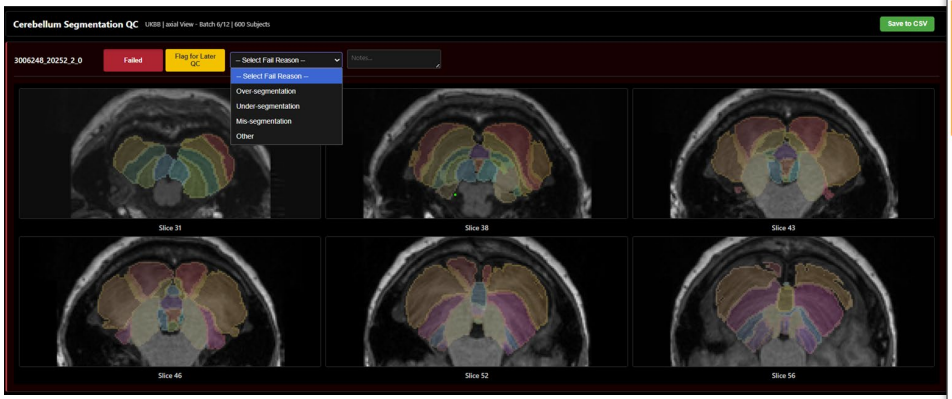
Fail BanksSTS + STG + MTG

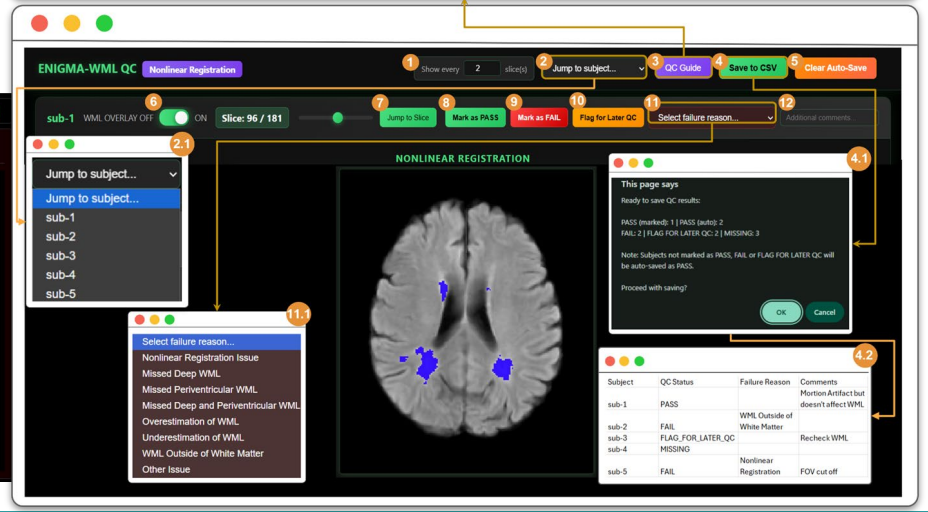


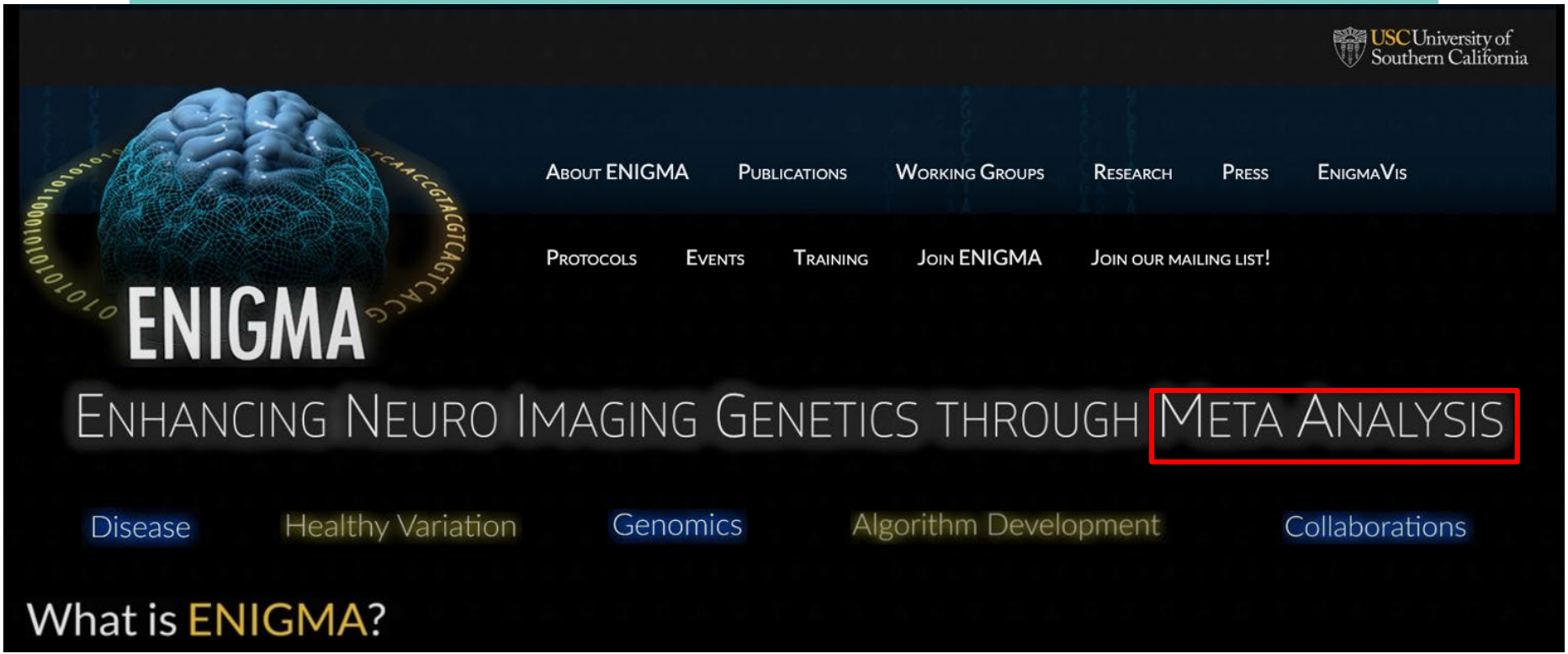
Sunanda Somu, MS









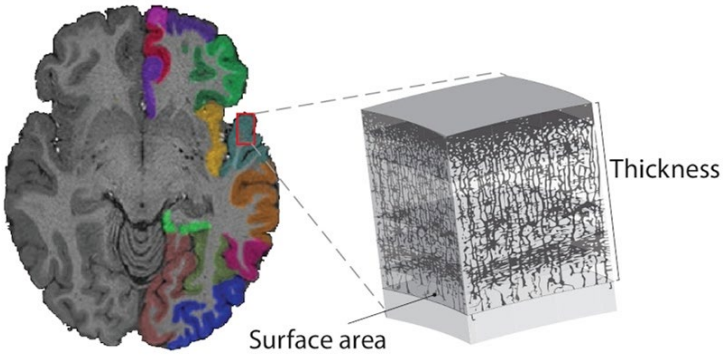


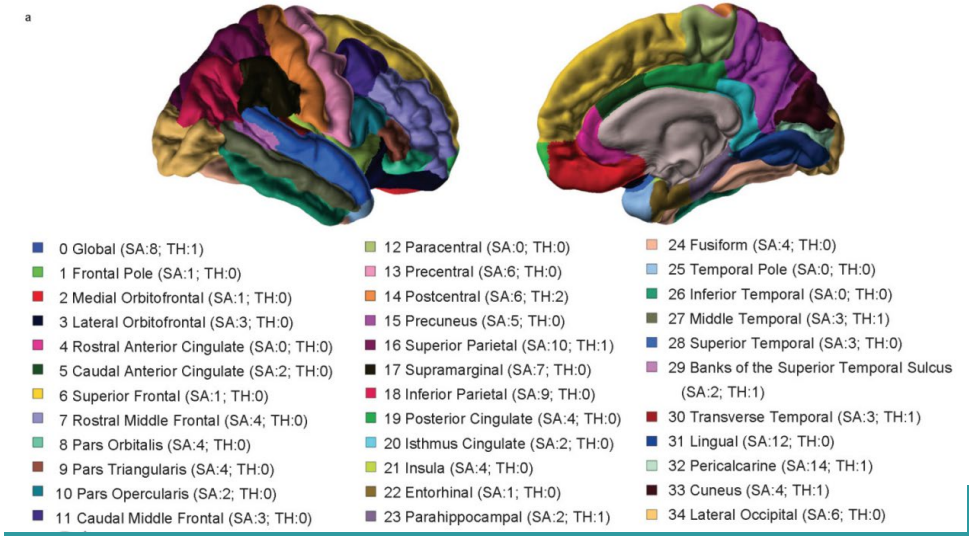
Distributed scientific collaboration. Site related heterogeneity weighted.
Effects are weighted by the variability in the site level data / or its sample size

One of ENIGMA's genome wide association studies:
The genetic architecture of the human cortex

Identifying genetic variants associated with normal variation in cortical surface area, thickness, and their regional patterning

Total sample size ~51,000 MRIs
Discovery sample size ~35,000 MRI

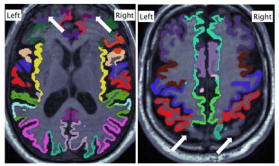




Grasby, Jahanshad et al Science 2020

Genome-Wide Meta Analysis:

Individual sites



Genetics Protocols

ENIGMA-git / ENIGMA

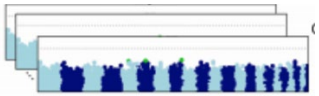
ENIGMA Consortium GWAS Protocols GWAS of the Cortex (ENIGMA3)

Version 1.1 – July 14, 2015

Written by Derrek Hibar and Neda Jahanshad

Please address any questions our Google group. Also, please check our FAQ to see if your question has already been answered.

We have tutorial videos available for running our scripts (run0, run1, run2) on our YouTube channel as well.



Imaging distribution statistics
Association summary statistics
No individual data

Meta-analysis + QC

For each SNP, across cohorts $i = 1...K$:

Each cohort contributes an effect estimate $\hat{\beta}_i$ (or Z-score) and its standard error SE_i from its own GWAS.

Inverse-variance weight for cohort i:

$$w_i = \frac{1}{SE_i^2}$$

Pooled (meta-analytic) effect estimate:

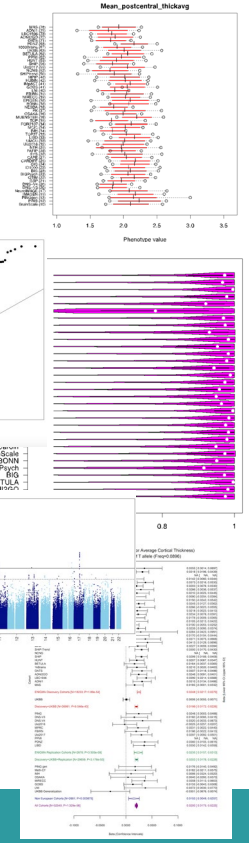
$$\hat{\beta}_{meta} = \frac{\sum_{i=1}^{K} w_i \hat{\beta}_i}{\sum_{i=1}^{K} w_i}$$

Standard error of the pooled estimate:

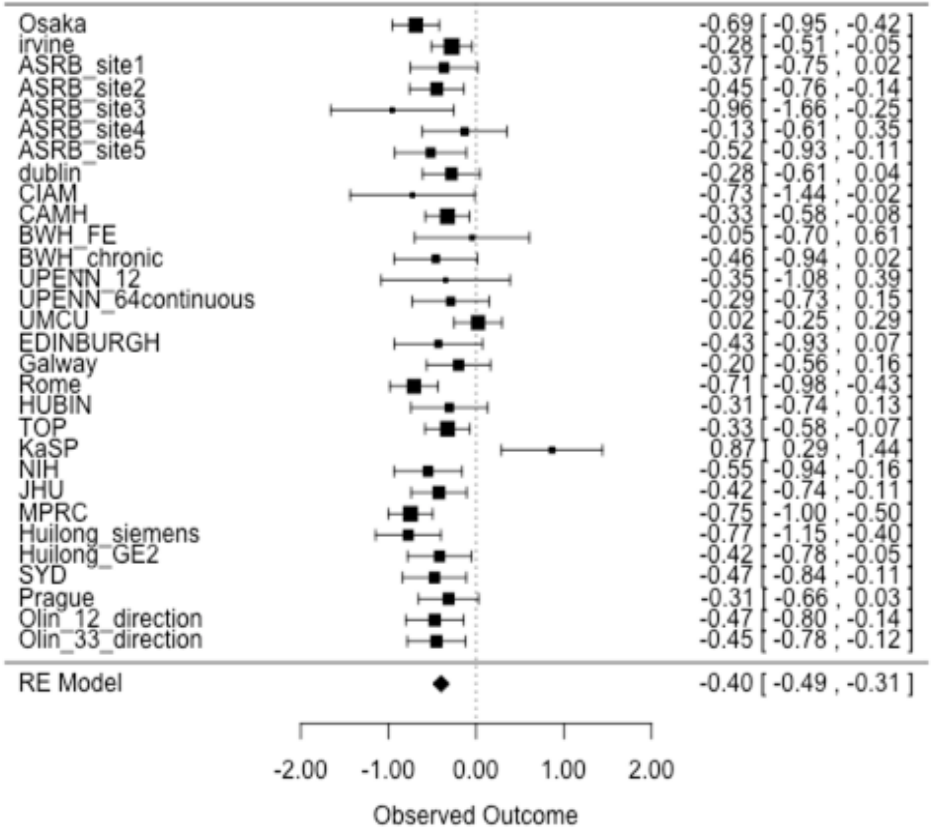
$$SE_{meta} = \sqrt{\frac{1}{\sum_{i=1}^{K} w_i}}$$

Meta-analytic Z-statistic and p-value:

$$Z_{meta} = \frac{\hat{\beta}_{meta}}{SE_{meta}}, \qquad p_{meta} = 2\left(1 - \Phi(|Z_{meta}|)\right)$$



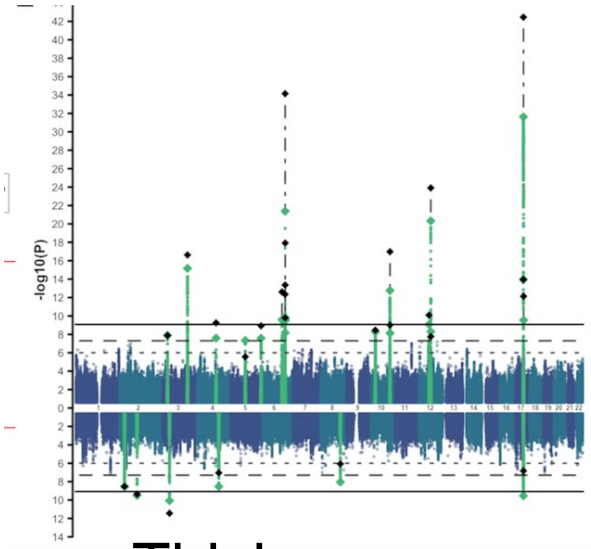
Error detection - Forest Plots



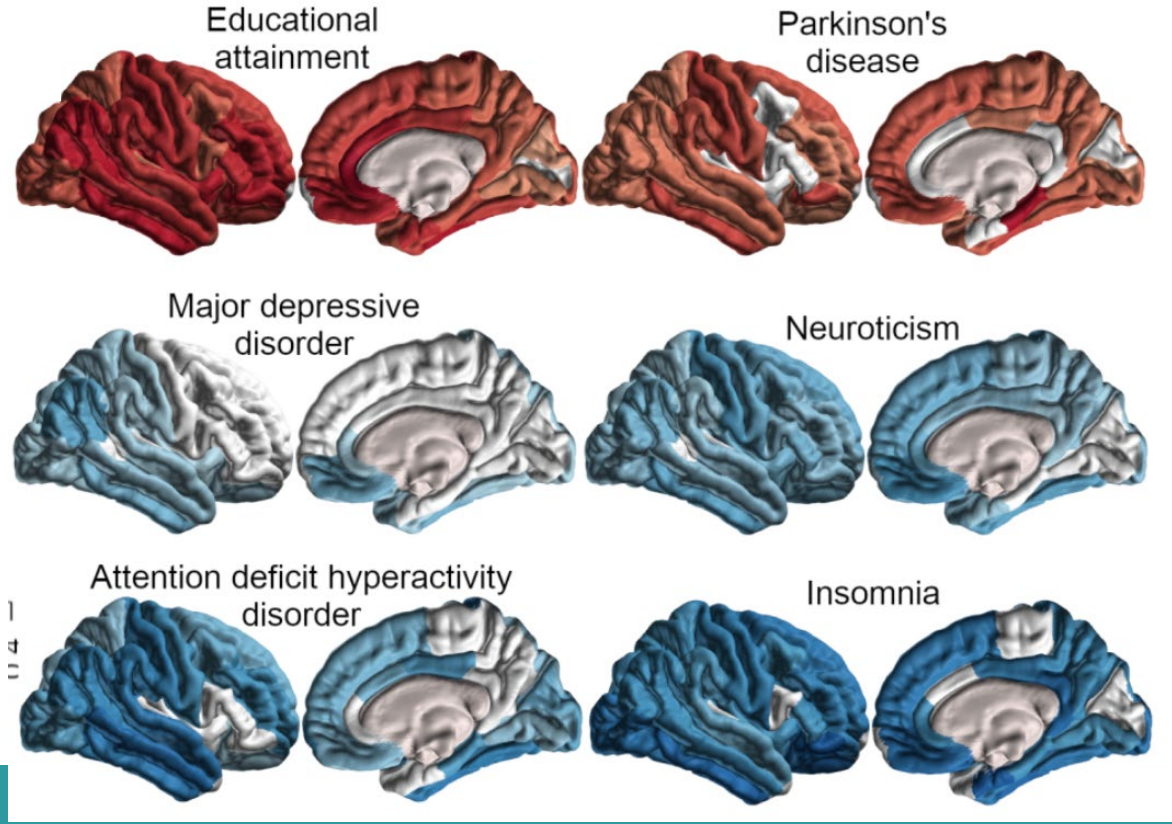
Novel insights into brain development + disease

200+ significant loci after replication ($p < 8.3 \times 10^{-10}$)

Surface Area



Thickness



Meta-Analysis

	Control	Cases
Study 1		
Study 2		
Study 3		
Study 4		

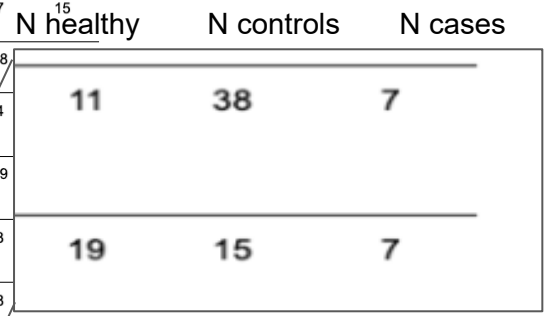
Imbalance across sites

	Control	Cases
Study 1		
Study 2		
Study 3		
Study 4		

	Control	Cases
Study 1		
Study 2		
Study 3		
Study 4		

Rare condition? Few cases?

BiDirect	-	48.6 (7.3)	49.5 (7.3)	-	56.3	66.9	0	316	124
CHU Montpellier BICS	42.2 (11.7)	45.3 (14.4)	44.2 (13.8)	71.4	86.4	90.0	21	22	20
CHU Montpellier IMPACT	-	38.3 (12.9)	38.6 (9.8)	-	94.4	85.7	0	18	7
CHU Montpellier UF8555	37.3 (7.9)	35.4 (8.4)	37.9 (9.5)	100.0	100.0	100.0	28	47	42
DCHS	30.9 (5.6)	29.7 (6.4)	32.3 (6.0)	100.0	100.0	100.0	22	20	9
MOODS team DEP-ARREST CLIN	35.5 (13.0)	33.2 (10.2)	34.8 (14.3)	67.7	70.8	63.0	31	24	27
ETPB-STB	33.9 (11.3)	34.3 (7.9)	34.8 (11.9)	65.0	50.0	63.6	20	12	11
Fondazione Santa Lucia _ Neuropsychiatry Lab / (FSL_NPL)-OCD sample	39.3 (10.5)	39.9 (11.1)	38.1 (7.6)	35.4	34.9	36.4	82	43	11
FSL_NPL - Schizophrenia sample	-	40.7 (11.7)	43.9 (9.9)	-	32.7	44.4	0	49	18
Ghent	33.6 (13.4)	38.2 (11.7)	34.8 (11.9)	94.1	53.3	69.2	17	15	13
GIPSI BD	-	31.5 (10.3)	34.4 (11.3)	-	25.0	25.0	0	24	8
GIPSI SZ	-	39.7 (11.9)	40.0 (11.4)	-	62.2	68.0	0	45	25
Grady Trauma Project Emory University	39.4 (13.0)	39.3 (12.1)	36.1 (10.9)	97.8	100.0	100.0	46	51	14
Halifax	40.4 (13.7)	42.8 (17.8)	51.0 (16.5)	64.3	41.2	57.1	42	17	7
Houston BD	24.1 (14.7)	29.1 (16.4)	30.2 (11.2)	43.1	48.4	64.3	51	64	14
LTS Colorado site 1	28.7 (1.1)	28.6 (0.9)	28.5 (0.5)	63.6	37.0	76.9	55	92	13
LTS Colorado site 2	28.6 (0.6)	28.7 (0.8)	28.6 (0.5)	73.3	62.5	66.7	30	72	9
McGill University	35.8 (8.0)	44.5 (11.3)	39.4 (10.6)	53.3	50.0	77.8	15	14	9
MR-IMPACT	-	15.0 (1.4)	15.2 (1.2)	-	67.1	91.4	0	70	35
PAFIP	-	29.8 (8.5)	27.3 (7.8)	-	36.4	26.7	0	77	15
San Raffaele Hospital	-	47.7 (10.9)	46.7 (10.2)	-	63.5	74.7	0	288	
Sydney Bipolar Kids and Siblings	20.9 (3.6)	23.0 (4.8)	24.1 (4.0)	52.2	65.5	75.0	67	84	
Sydney Brain and Mind Centre	-	19.5 (3.5)	18.1 (2.6)	-	65.1	68.8	0	169	
UCSF Adolescent MDD	15.4 (1.4)	15.7 (1.3)	16.0 (1.1)	60.0	55.8	71.4	40	43	
University of Minnesota Adolescent MDD	16.5 (2.2)	15.5 (1.8)	16.9 (2.1)	72.7	78.9	71.4	11	38	
University of Texas- Austin -Bipolar Seed Program	20.5 (1.7)	21.0 (2.3)	22.4 (2.0)	68.4	73.3	57.1	19	15	7
University of Washington/ Harvard	12.0 (1.8)	12.4 (2.6)	14.4 (1.8)	46.7	35.5	60.0	45	31	5
Yale School of Medicine	-	19.3 (3.4)	19.0 (3.2)	-	60.4	73.6	0	111	53
Total	**29.6 (13.1)**	**34.0 (14.8)**	**35.4 (14.7)**	**62.9**	**59.8**	**71.8**	**642**	**1871**	**652**



Van Velzen et al 2025, Translational Psychiatry

Mega-Analysis Motivation Case - Guess

1) 1 death every 11 minutes / over 700.000 deaths a year worldwide
2) In 2019 alone, estimated costs IN THE US were nearly **$94 BILLION**
3) 75 % of cases occurring in low- and middle-income countries
4) Over half of all deaths occur before age 50
5) One of the 10 leading cause of death among all age groups 10-64
6) 8.9% of high school students were affected in the US

Statistics from: America's Health Rankings analysis of CDC WONDER, Multiple Cause of Death Files, United Health Foundation, AmericasHealthRankings.org, Accessed 2022.
DOI: 10.1016/j.psychres.2025.116597
https://www.who.int/publications/i/item/9789240026643

Mega-Analysis - Guess

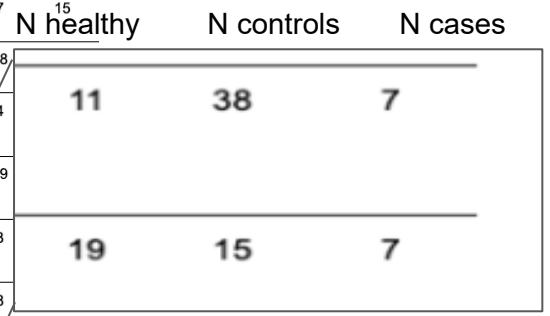
1) 1 death every 11 minutes / over 700.000 deaths a year worldwide
2) In 2019 alone, estimated costs IN THE US were nearly **$94 BILLION**
3) 75 % of cases occurring in low- and middle-income countries
4) Over half of all deaths occur before age 50
5) One of the 10 leading cause of death among age groups 10-64
6) 8.9% of high school students attempt suicide in the US

Suicide Attempt

Statistics from: America's Health Rankings analysis of CDC WONDER, Multiple Cause of Death Files, United Health Foundation, AmericasHealthRankings.org, Accessed 2022.
DOI: 10.1016/j.psychres.2025.116597
https://www.who.int/publications/i/item/9789240026643

Rare condition? Few cases?

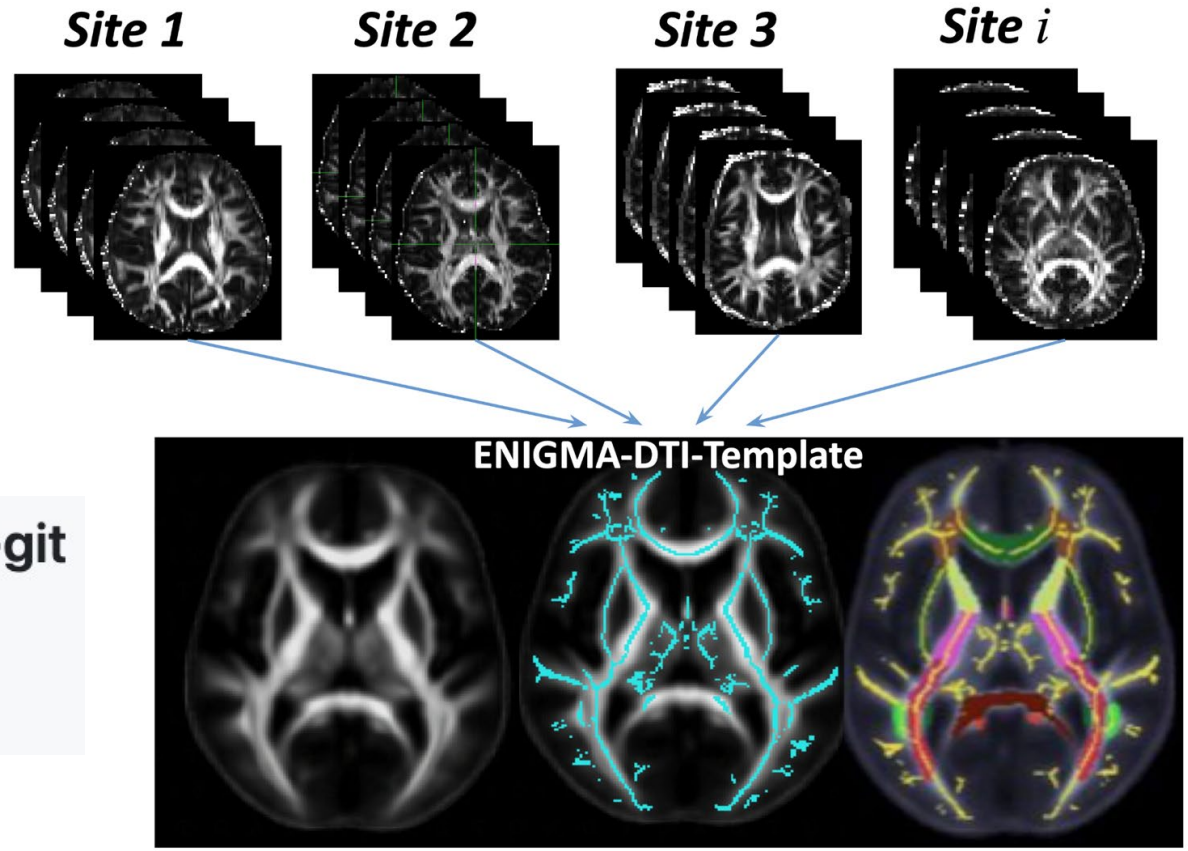
BiDirect	-	48.6 (7.3)	49.5 (7.3)	-	56.3	66.9	0	316	124
CHU Montpellier BICS	42.2 (11.7)	45.3 (14.4)	44.2 (13.8)	71.4	86.4	90.0	21	22	20
CHU Montpellier IMPACT	-	38.3 (12.9)	38.6 (9.8)	-	94.4	85.7	0	18	7
CHU Montpellier UF8555	37.3 (7.9)	35.4 (8.4)	37.9 (9.5)	100.0	100.0	100.0	28	47	42
DCHS	30.9 (5.6)	29.7 (6.4)	32.3 (6.0)	100.0	100.0	100.0	22	20	9
MOODS team DEP-ARREST CLIN	35.5 (13.0)	33.2 (10.2)	34.8 (14.3)	67.7	70.8	63.0	31	24	27
ETPB-STB	33.9 (11.3)	34.3 (7.9)	34.8 (11.9)	65.0	50.0	63.6	20	12	11
Fondazione Santa Lucia _ Neuropsychiatry Lab / (FSL_NPL)-OCD sample	39.3 (10.5)	39.9 (11.1)	38.1 (7.6)	35.4	34.9	36.4	82	43	11
FSL_NPL - Schizophrenia sample	-	40.7 (11.7)	43.9 (9.9)	-	32.7	44.4	0	49	18
Ghent	33.6 (13.4)	38.2 (11.7)	34.8 (11.9)	94.1	53.3	69.2	17	15	13
GIPSI BD	-	31.5 (10.3)	34.4 (11.3)	-	25.0	25.0	0	24	8
GIPSI SZ	-	39.7 (11.9)	40.0 (11.4)	-	62.2	68.0	0	45	25
Grady Trauma Project Emory University	39.4 (13.0)	39.3 (12.1)	36.1 (10.9)	97.8	100.0	100.0	46	51	14
Halifax	40.4 (13.7)	42.8 (17.8)	51.0 (16.5)	64.3	41.2	57.1	42	17	7
Houston BD	24.1 (14.7)	29.1 (16.4)	30.2 (11.2)	43.1	48.4	64.3	51	64	14
LTS Colorado site 1	28.7 (1.1)	28.6 (0.9)	28.5 (0.5)	63.6	37.0	76.9	55	92	13
LTS Colorado site 2	28.6 (0.6)	28.7 (0.8)	28.6 (0.5)	73.3	62.5	66.7	30	72	9
McGill University	35.8 (8.0)	44.5 (11.3)	39.4 (10.6)	53.3	50.0	77.8	15	14	9
MR-IMPACT	-	15.0 (1.4)	15.2 (1.2)	-	67.1	91.4	0	70	35
PAFIP	-	29.8 (8.5)	27.3 (7.8)	-	36.4	26.7	0	77	15
San Raffaele Hospital	-	47.7 (10.9)	46.7 (10.2)	-	63.5	74.7	0	288	
Sydney Bipolar Kids and Siblings	20.9 (3.6)	23.0 (4.8)	24.1 (4.0)	52.2	65.5	75.0	67	84	
Sydney Brain and Mind Centre	-	19.5 (3.5)	18.1 (2.6)	-	65.1	68.8	0	169	
UCSF Adolescent MDD	15.4 (1.4)	15.7 (1.3)	16.0 (1.1)	60.0	55.8	71.4	40	43	
University of Minnesota Adolescent MDD	16.5 (2.2)	15.5 (1.8)	16.9 (2.1)	72.7	78.9	71.4	11	38	
University of Texas- Austin -Bipolar Seed Program	20.5 (1.7)	21.0 (2.3)	22.4 (2.0)	68.4	73.3	57.1	19	15	7
University of Washington/ Harvard	12.0 (1.8)	12.4 (2.6)	14.4 (1.8)	46.7	35.5	60.0	45	31	5
Yale School of Medicine	-	19.3 (3.4)	19.0 (3.2)	-	60.4	73.6	0	111	53
Total	**29.6 (13.1)**	**34.0 (14.8)**	**35.4 (14.7)**	**62.9**	**59.8**	**71.8**	**642**	**1871**	**652**



Van Velzen et al 2025, Translational Psychiatry

Standardization of dMRI Imaging Methods

White matter microstructure



ENIGMA-git

Jahanshad et al 2013;
Builds on FSL's TBSS (Smith et al 2006)

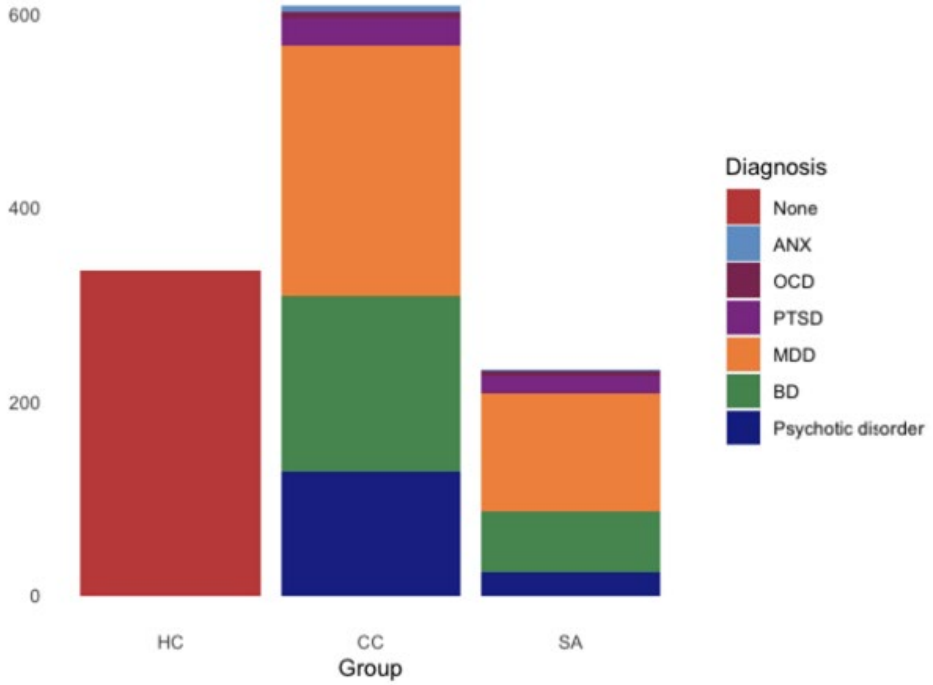
Power to conduct first of a kind study

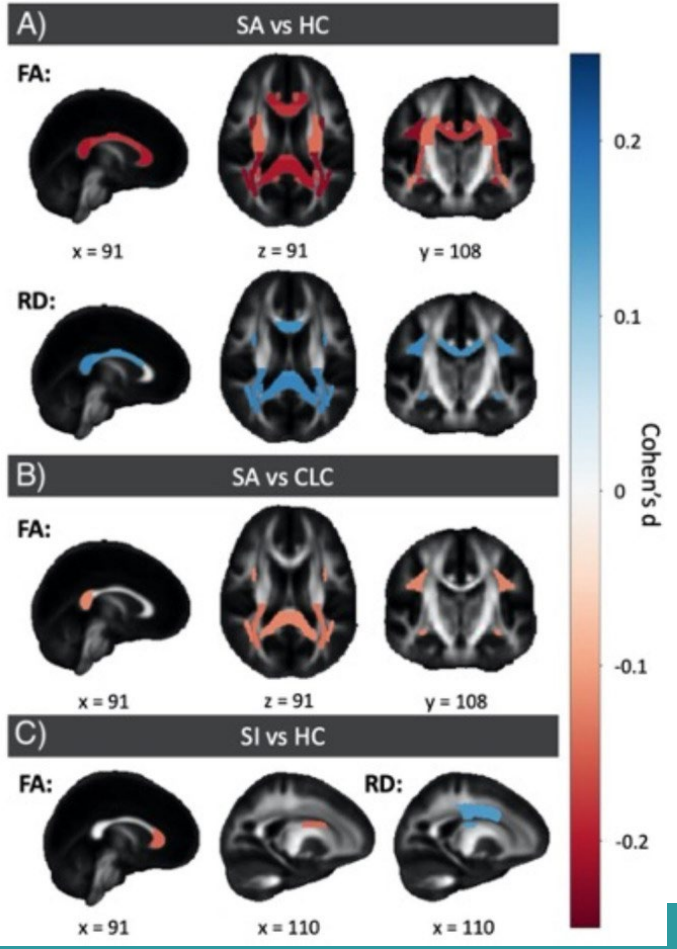
Site	Age HC	Age CLC	Age suicide attempt group	% female HC	% female CLC	% female suicide attempt group	Total N HC	Total N CLC	Total N suicide attempt group
BiDirect	-	48.6 (7.3)	49.5 (7.3)	-	56.3	66.9	0	316	124
CHU Montpellier BICS	42.2 (11.7)	45.3 (14.4)	44.2 (13.8)	71.4	86.4	90.0	21	22	20
CHU Montpellier IMPACT	-	38.3 (12.9)	38.6 (9.8)	-	94.4	85.7	0	18	7
CHU Montpellier UF8555	37.3 (7.9)	35.4 (8.4)	37.9 (9.5)	100.0	100.0	100.0	28	47	42
DCHS	30.9 (5.6)	29.7 (6.4)	32.3 (6.0)	100.0	100.0	100.0	22	20	9
MOODS team DEP-ARREST CLIN	35.5 (13.0)	33.2 (10.2)	34.8 (14.3)	67.7	70.8	63.0	31	24	27
ETPB-STB	33.9 (11.3)	34.3 (7.9)	34.8 (11.9)	65.0	50.0	63.6	20	12	11
Fondazione Santa Lucia _ Neuropsychiatry Lab / (FSL_NPL)-OCD sample	39.3 (10.5)	39.9 (11.1)	38.1 (7.6)	35.4	34.9	36.4	82	43	11
FSL_NPL - Schizophrenia sample	-	40.7 (11.7)	43.9 (9.9)	-	32.7	44.4	0	49	18
Ghent	33.6 (13.4)	38.2 (11.7)	34.8 (11.9)	94.1	53.3	69.2	17	15	13
GIPSI BD	-	31.5 (10.3)	34.4 (11.3)	-	25.0	25.0	0	24	8
GIPSI SZ	-	39.7 (11.9)	40.0 (11.4)	-	62.2	68.0	0	45	25
Grady Trauma Project Emory University	39.4 (13.0)	39.3 (12.1)	36.1 (10.9)	97.8	100.0	100.0	46	51	14
Halifax	40.4 (13.7)	42.8 (17.8)	51.0 (16.5)	64.3	41.2	57.1	42	17	7
Houston BD	24.1 (14.7)	29.1 (16.4)	30.2 (11.2)	43.1	48.4	64.3	51	64	14
LTS Colorado site 1	28.7 (1.1)	28.6 (0.9)	28.5 (0.5)	63.6	37.0	76.9	55	92	13
LTS Colorado site 2	28.6 (0.6)	28.7 (0.8)	28.6 (0.5)	73.3	62.5	66.7	30	72	9
McGill University	35.8 (8.0)	44.5 (11.3)	39.4 (10.6)	53.3	50.0	77.8	15	14	9
MR-IMPACT	-	15.0 (1.4)	15.2 (1.2)	-	67.1	91.4	0	70	35
PAFIP	-	29.8 (8.5)	27.3 (7.8)	-	36.4	26.7	0	77	15
San Raffaele Hospital	-	47.7 (10.9)	46.7 (10.2)	-	63.5	74.7	0	288	79
Sydney Bipolar Kids and Siblings	20.9 (3.6)	23.0 (4.8)	24.1 (4.0)	52.2	65.5	75.0	67	84	24
Sydney Brain and Mind Centre	-	19.5 (3.5)	18.1 (2.6)	-	65.1	68.8	0	169	32
UCSF Adolescent MDD	15.4 (1.4)	15.7 (1.3)	16.0 (1.1)	60.0	55.8	71.4	40	43	14
University of Minnesota Adolescent MDD	16.5 (2.2)	15.5 (1.8)	16.9 (2.1)	72.7	78.9	71.4	11	[illegible]	[illegible]
University of Texas- Austin -Bipolar Seed Program	20.5 (1.7)	21.0 (2.3)	22.4 (2.0)	68.4	73.3	57.1	19	[illegible]	[illegible]
University of Washington/ Harvard	12.0 (1.8)	12.4 (2.6)	14.4 (1.8)	46.7	35.5	60.0	45	31	5
Yale School of Medicine	-	19.3 (3.4)	19.0 (3.2)	-	60.4	73.6	0	111	53
Total	**29.6 (13.1)**	**34.0 (14.8)**	**35.4 (14.7)**	**62.9**	**59.8**	**71.8**	**642**	**1871**	**652**

N healthy	N controls	N cases
642	1871	652

Van Velzen et al 2025, Translational Psychiatry

Van Velzen et al 2025,
Translational Psychiatry





Many exciting developments in statistical data harmonization

README.md

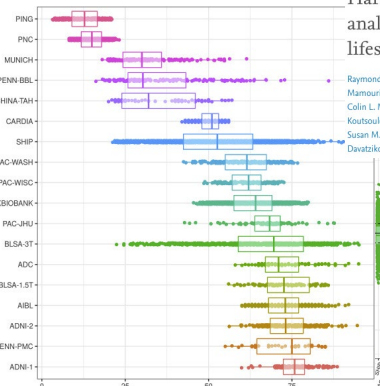
ComBatHarmonization

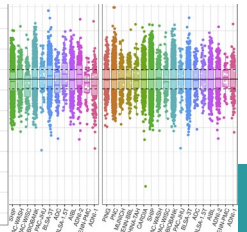
Harmonization of multi-site imaging data with ComBat

Jean-Philippe Fortin, Drew Parker, Birkan Tunc, Takanori Watanabe, Ma
Elliott, Kosha Ruparel, David R Roalf, Theodore D Satterthwaite, Ruben
Gur, Raquel E Gur, Robert T Schultz, Ragini Verma, Russell T Shinohara.
Harmonization Of Multi-Site Diffusion Tensor Imaging Data.
NeuroImage, 161, 149-170, 2017

Jean-Philippe Fortin, Nicholas Cullen, Yvette I. Sheline, Warren D. Taylo
Irem Aselcioglu, Philip A. Cook, Phil Adams, Crystal Cooper, Maurizio Fa
Patrick J. McGrath, Melvin McInnis, Mary L. Phillips, Madhukar H. Trived
Myrna M. Weissman, Russell T. Shinohara. **Harmonization of cortical thickness measurements across scanners and sites.** NeuroImage, 167
104-120, 2018

W. Evan Johnson and Cheng Li, **Adjusting batch effects in microarray expression data using empirical Bayes m**
127, 2007.





ELSEVIER

NeuroImage
Volume 220, 15 October 2020, 117129

Longitudinal ComBat: A method for harmonizing longitudinal multi-scanner imaging data ☆

Cook [c, d], Christos Davatzikos [e], Yvette I. Sheline [c, d, f], Russell
ner's Disease Neuroimaging Initiative

NEUROIMAGE ELSEVIER FREE Full-Text Article

Neuroimage. 2020 Oct 15; 220: 117127. PMCID: PMC7573655
doi: 10.1016/j.neuroimage.2020.117127 PMID: 32634595

Neuroharmony: A new tool for harmonizing volumetric MRI data from unseen scanners

Rafael Garcia-Dias,[a,*] Cristina Scarpazza,[a,b] Lea Baecker,[a] Sandra Vieira,[a] Walter H.L. Pinaya,[a,c] Aiden Corvin,[d] Alberto Redolfi,[e] Barnaby Nelson,[f,g] Benedicto Crespo-Facorro,[h,i,j,k] Colm McDonald,[l] Diana Tordesillas-Gutiérrez,[h,m] Dara Cannon,[l] David Mothersill,[n] Dennis Hernaus,[o] Derek Morris,[p] Esther Setien-Suero,[h,k] Gary Donohoe,[n] Giovanni Frisoni,[q,r] Giulia Tronchin,[l] João Sato,[c] Machteld Marcelis,[o] Haren,[t] Oliver Gruber,[u,v] Patrick McGorry,[f,g] Paul Amminger,[f,g] René S. Kahn,[z,aa] Rosa Ayesa-Arriola,[h,k] Therese van Amelsvoort,[o] e Calhoun,[ab,s] Wiepke Cahn,[z] and Andrea Mechelli[a]

ELSEVIER

NeuroImage
Volume 208, March 2020, 116450

Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan

Raymond Pomponio [a], Guray Erus [a], Mohamad Habes [a, b], Jimit Doshi [a], Dhivya Srinivasan [a], Elizabeth Mamourian [a], Vishnu Bashyam [a], Ilya M. Nasrallah [a, 8], Theodore D. Satterthwaite [], Yong Fan [a], Lenore J. Launer [c], Colin L. Masters [d], Paul Maruff [d], Chuanjun Zhuo [e, f], Henry Völzke [h], Sterling C. Johnson [], Jurgen Fripp [], Nikolaos Koutsouleris [k], Daniel H. Wolf [], Raquel Gur [g,], Ruben Gur [g,], John Morris [m], Marilyn S. Albert [n], Hans J. Grabe [o], Susan M. Resnick [p], R. Nick Bryan [q], David A. Wolk [], Russell T. Shinohara [a, r,], Haochang Shou [a, r,], Christos Davatzikos [a, ,]

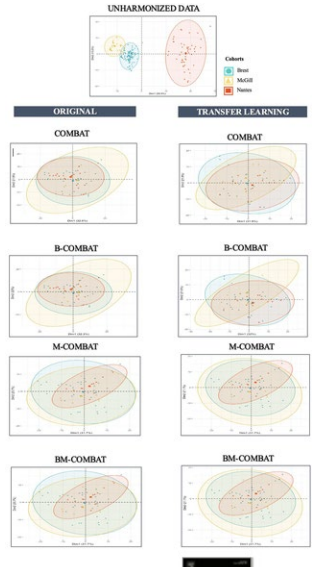


Accommodating site variation in neuroimaging data using normative and hierarchical Bayesian models

Johanna M. M. Bayer, Richard Dinga, Seyed Mostafa Kia, Akhil R. Kottaram, Thomas Wolfers, Jinglei Lv, Andrew Zalesky, Lianne Schmaal, Andre Marquand

doi: https://doi.org/10.1101/2021.02.09.430363

This article is a preprint and has not been certified by peer review [what does this mean?].



ELSEVIER

NeuroImage
Volume 245, 15 December 2021, 118703

PLOS ONE

OPEN ACCESS PEER-REVIEWED
RESEARCH ARTICLE

A transfer learning approach to facilitate ComBat-based harmonization of multicentre radiomic features in new datasets

Ronrick Da-ano, François Lucia, Ingrid Masson, Ronan Abgral, Joanne Alfieri, Caroline Rousseau, Augustin Mervoyer, Caroline Reinhold, Olivier Pradier, Ulrike Schick, Dimitris Visvikis, Mathieu Hatt

A multi-scanner neuroimaging data harmonization using RAVEL and ComBat

Mahbaneh Eshaghzadeh Torbati [a], Davneet S. Minhas [b], Ghasan Ahmad [c], Erin E. O'Connor [c], John Muschelli [d], Charles M. Laymon [b], Zixi Yang [b], Ann D. Cohen [e], Howard J. Aizenstein [e], William E. Klunk [e], Bradley T. Christian [f], Seong Jae Hwang [a, g], Ciprian M. Crainiceanu [d], Dana L. Tudorascu [e, h]

ComBat for Harmonization in Multi-Site MRI Studies

ComBat is a statistical harmonization method that uses an Empirical Bayes approach to estimate **site-specific shifts and scaling factors**

- **Johnson et al 2007; Fortin et al 2017**

ComBat[14]**:** ComBat is a location and scale adjustment model where, for site i and individual j, it assumes that the feature measurements are modeled as a linear combination of site effects and non-site effects, written as:

$$y_{ijv} = \alpha_v + X_{ij}^T \beta_v + \gamma_{iv} + \delta_{iv} e_{ijv}$$

where α_v is the overall mean per feature v, β_v is the vector of corresponding coefficients to the covariate matrix X, γ_{iv} is an offset from the grand mean per site i and feature v, e_{ijv} is the residual vector, and δ_{iv} is the multiplicative site effect of site i on feature v. ComBat removes the additive and multiplicative effects from the residuals using

$$e_{ijv}^{ComBat} = \frac{y_{ijv} - \widehat{\alpha_v} - X_{ij}^T \widehat{\beta_v} - \gamma_{iv}^*}{\delta_{iv}^*}$$

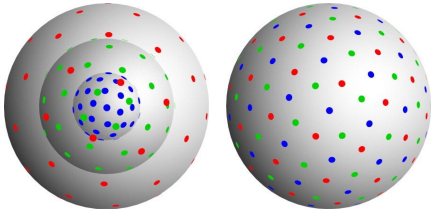
where γ_{iv}^* and δ_{iv}^* are estimated using an empirical Bayes framework. The harmonized data are then calculated by

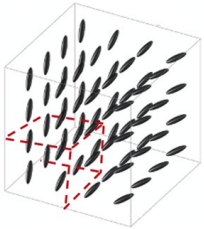
$$y_{ijv}^{ComBat} = e_{ijv}^{ComBat} + \widehat{\alpha_v} + X_{ij}^T \widehat{\beta_v}$$

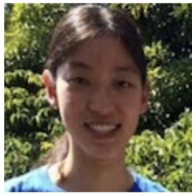
where $\widehat{\alpha_v}$ and $\widehat{\beta_v}$ are the respective parameter estimates.

https://github.com/andy1764/ComBatFamily https://github.com/PennSIVE/ComBatFam_Pipeline

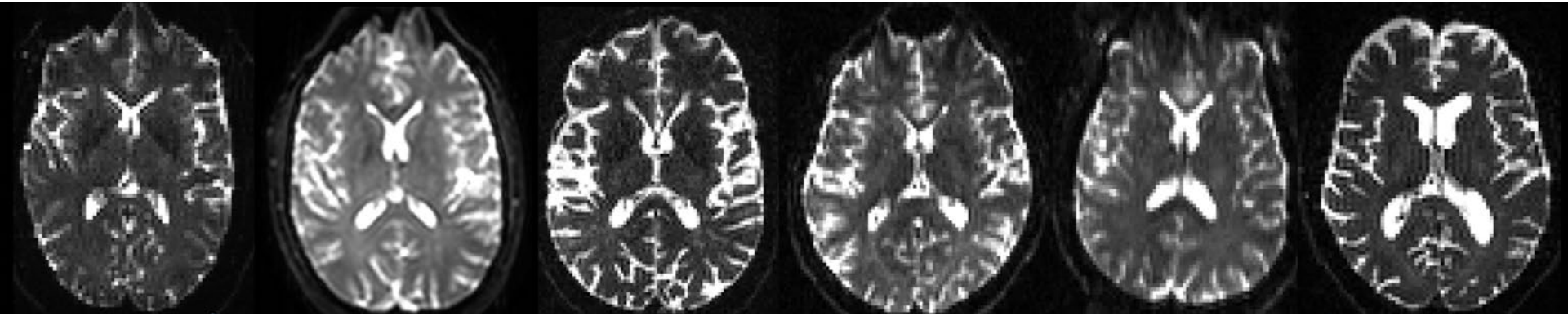
Diffusion MRI protocols are particularly heterogeneous





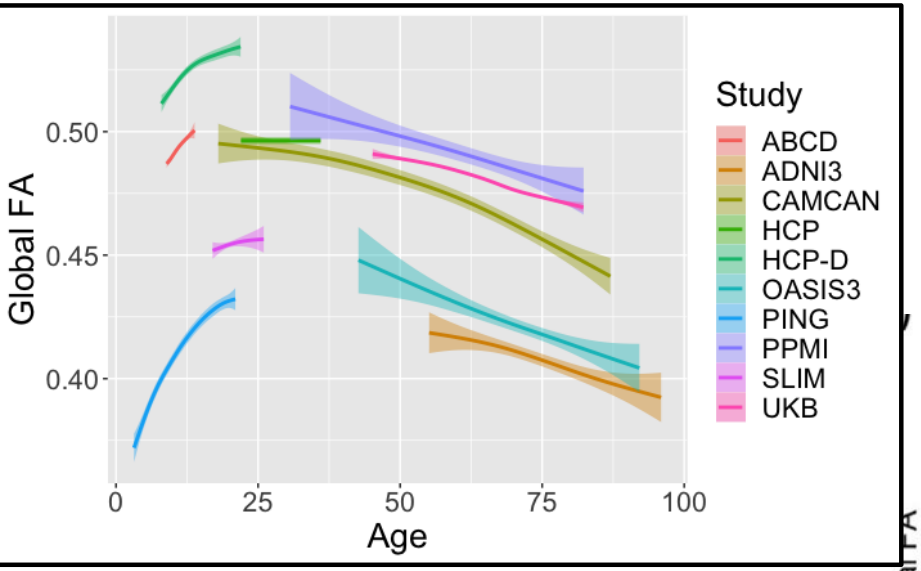


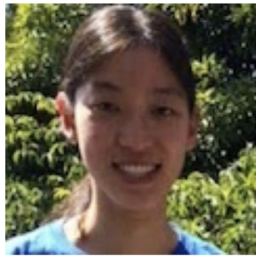
Alyssa Zhu, PhD



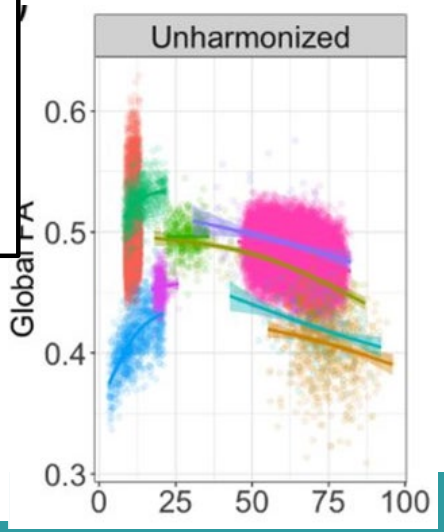
Zhu et al 2018, Nir et al 2018

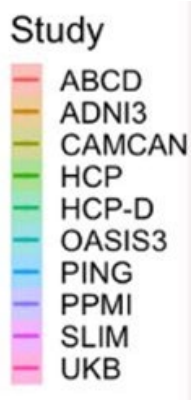
Building a template for regional lifespan trajectories





Alyssa Zhu, PhD

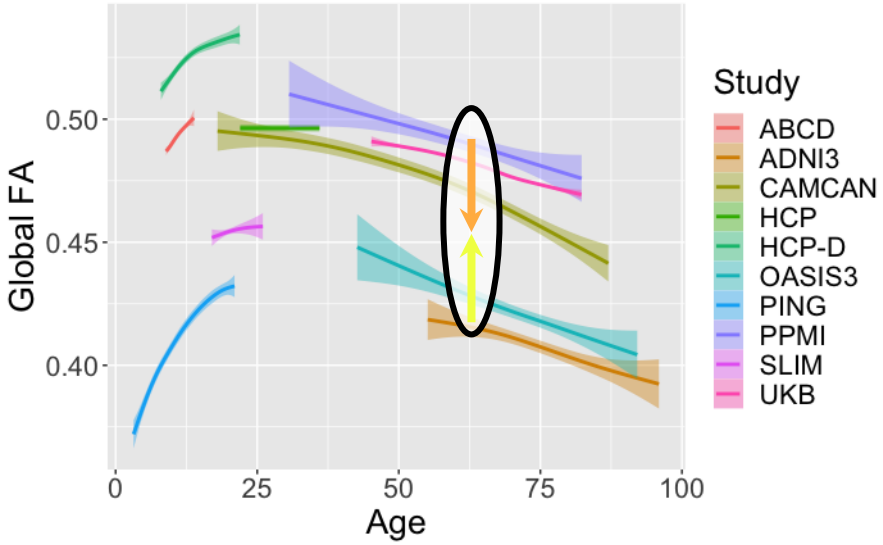


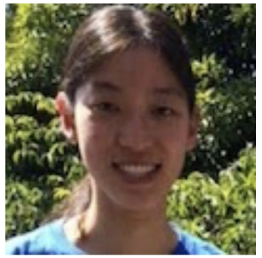


Zhu et al - eHarmonize - Scientific Data 2025

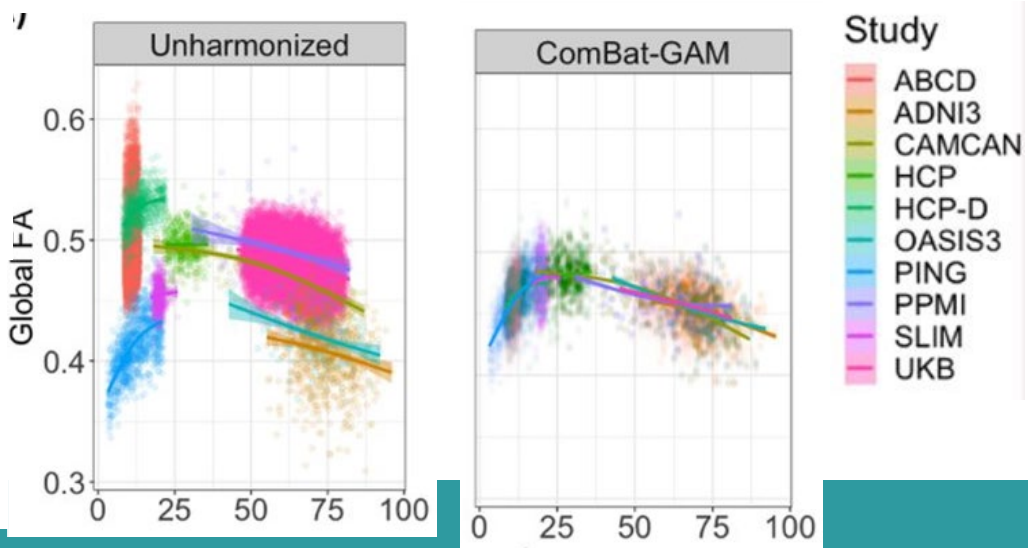
https://github.com/USC-LoBeS/eharmonize

Building a template for regional lifespan trajectories





Alyssa Zhu, PhD



Zhu et al - eHarmonize - Scientific Data 2025

https://github.com/USC-LoBeS/eharmonize

Building a template for regional lifespan trajectories

dMRI Studies and Subjects

MRI Harmonization

- **N = 40,898** (10 studies)
- Age Range: 3-95 years
- Healthy controls

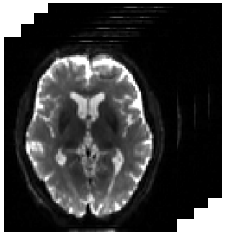
In multi-shell acquisitions, only the shells used for DTI processing were included

Study	Age (yrs)	N (%Male)	Scanner (3T)	*b*-values (volumes)	Resolution (mm^3)
PING	3-21	906 (52.2)	Siemens GE Philips	0 (4), 1000 (32) 0 (2), 1000 (30) 0 (3), 1000 (30)	2.5 x 2.5 x 2.5
HCP-D	8-21	617 (48.5)	Siemens	0 (28), 1500 (93)	1.5 x 1.5 x 1.5
ABCD	9-13	5611 (52.7)	Siemens	0 (7), 1000 (15)	1.7 x 1.7 x 1.7
SLIM	17-26	432 (45.1)	Siemens	0, 1000 (30)	2 x 2 x 2
HCP	22-36	363 (45.5)	Siemens	0 (18), 1000 (90)	1.25 x 1.25 x 1.25
Cam-CAN	18-87	305 (35.1)	Siemens	0 (3), 1000 (30)	2 x 2 x 2
PPMI	30-81	83 (63.9)	Siemens	0 (1), 1000 (64)	1.98 x 1.98 x 2
[illegible]k	45-82	31,986 (46.8)	Siemens	0 (8), 1000 (50)	2.5 x 2.5 x 2.5
	42-92	225 (48.4)	Siemens	0 (1), 1000 (64)	2.5 x 2.5 x 2.5
	55-95	370 (41.2)	Siemens GE Philips	0 (1-7), 1000 (30-64)	2 x 2 x 2

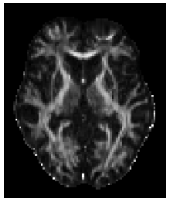
```
[17:42]c2025:03_python_endharm_test[edti_harm_test]$ eharmonize harmonize-fa --help
Usage: eharmonize harmonize-fa [OPTIONS]

Options:
  --incsv CSV           File path of the CSV file with the necessary covariates
                        and FA measures
  --outdir DIR          Directory to write out outputs to
  --reference version   (optional) Version number of desired reference
                        [default: v0.0]
  --rerun               (optional) if used, will overwrite previous outputs
  -h, --help            Show this message and exit.
[17:43]c2025:03_python_endharm_test[edti_harm_test]$ eharmonize apply-harmonization --help
Usage: eharmonize apply-harmonization [OPTIONS]

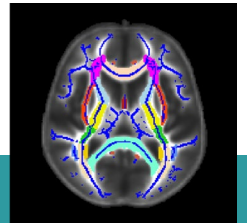
Options:
  --incsv CSV       File path of the CSV file with the necessary covariates and
                    FA measures
  --model MODEL     Model file as output by initial harmonization
  --log JSON        JSON file from initial harmonization
  --outdir DIR      Directory to write out outputs to
  --metric METRIC   (optional)  [default: FA]
  -h, --help        Show this message and exit.
```



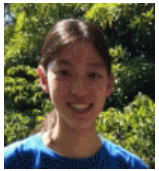
Model DTI



ENIGMA-DTI

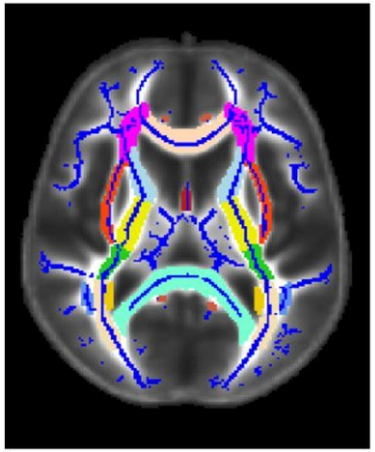


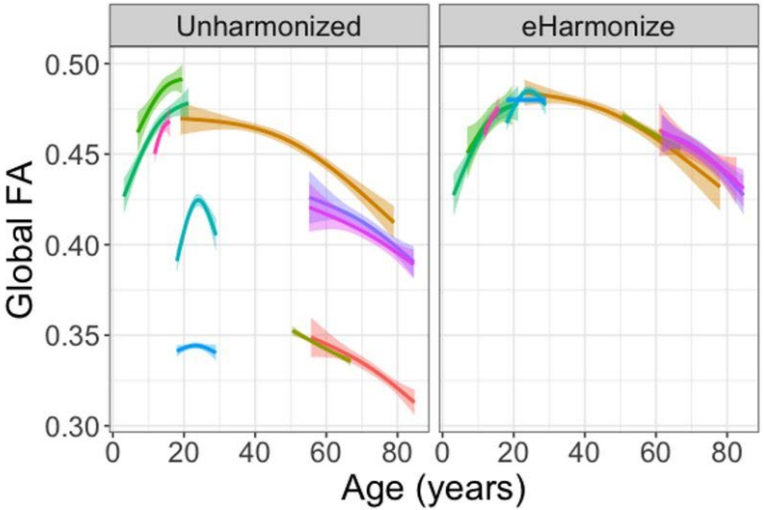
Harmonizing diffusion metrics across the lifespan to assess APOE4 affects

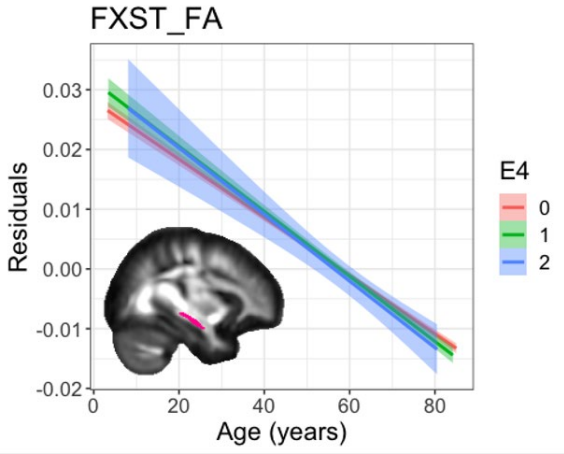


Alyssa Zhu

Normative lifespan curves for diffusion MRI metrics



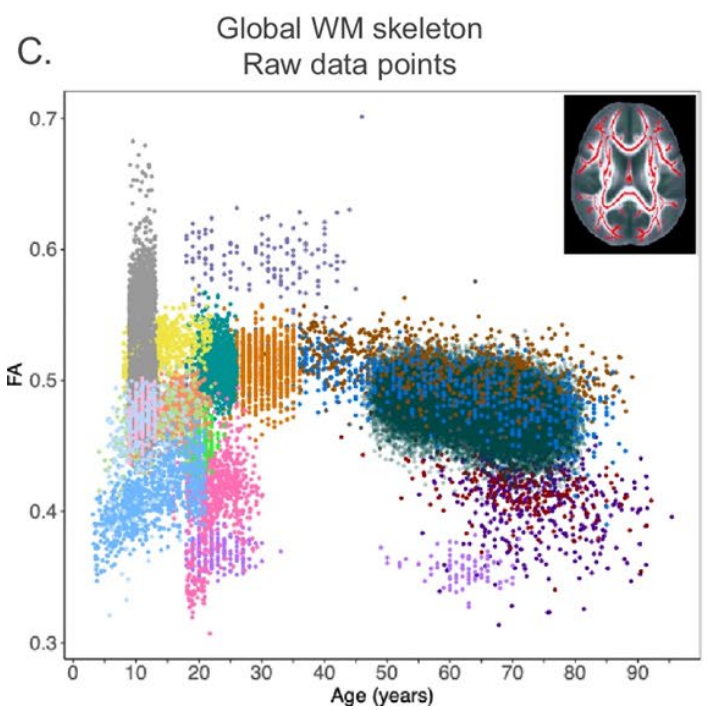


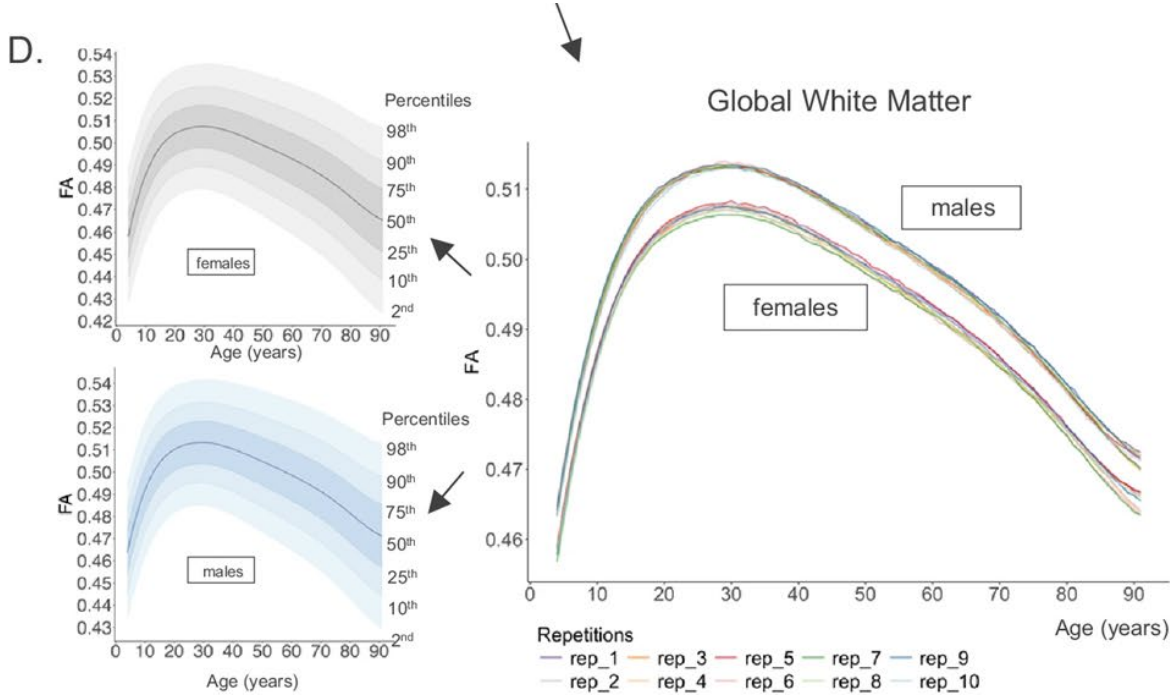


Antagonistic pleiotropy

Zhu et al - eHarmonize - Scientific Data 2025
https://github.com/USC-LoBeS/eharmonize

Expansion with Hierarchical Bayes Regression





Many exciting developments in statistical data harmonization

README.md

ComBatHarmonization

Harmonization of multi-site imaging data with ComBat

Jean-Philippe Fortin, Drew Parker, Birkan Tunc, Takanori Watanabe, Ma
Elliott, Kosha Ruparel, David R Roalf, Theodore D Satterthwaite, Ruben
Gur, Raquel E Gur, Robert T Schultz, Ragini Verma, Russell T Shinohara.
Harmonization Of Multi-Site Diffusion Tensor Imaging Data.
NeuroImage, 161, 149-170, 2017

Jean-Philippe Fortin, Nicholas Cullen, Yvette I. Sheline, Warren D. Taylo
Irem Aselcioglu, Philip A. Cook, Phil Adams, Crystal Cooper, Maurizio Fa
Patrick J. McGrath, Melvin McInnis
Myrna M. Weissman, Russell T. Shi
thickness measurements across
104-120, 2018

W. Evan Johnson and Cheng Li, Ad
expression data using empirical B
127, 2007.

ELSEVIER

NeuroImage
Volume 220, 15 October 2020, 117129

Longitudinal ComBat: A method for harmonizing
nner imaging data ☆

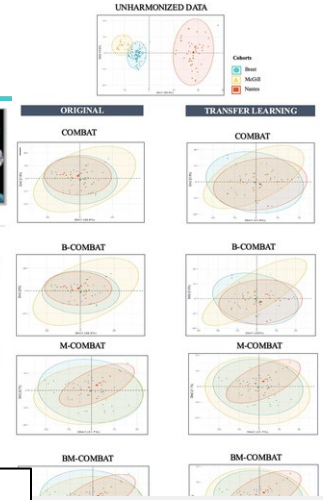
Cook [c, d], Christos Davatzikos [e], Yvette I. Sheline [c, d, f], Russell
ner's Disease Neuroimaging Initiative

NEUROIMAGE ELSEVIER FREE Full-Text Article

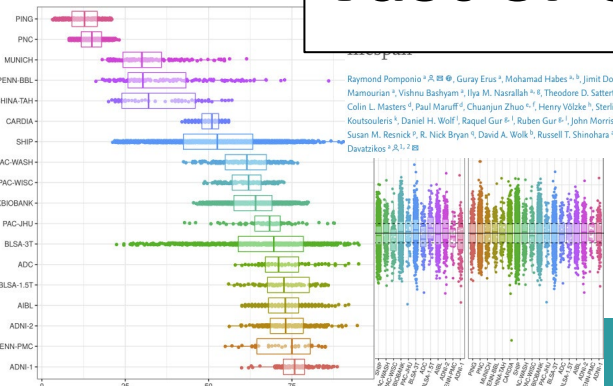
Neuroimage. 2020 Oct 15; 220: 117127.
doi: 10.1016/j.neuroimage.2020.117127

PMCID: PMC7573655
PMID: 32634595

Neuroharmony: A new tool for harmonizing volumetric MRI data from



What if site differences are too vast or confounded?



Raymond Pomponio [a], Guray Erus [a], Mohamad Habes [a, b], Jimit Doshi [a], Dhivya Srinivasan [a], Elizabeth Mamourian [a], Vishnu Bashyam [a], Ilya M. Nasrallah [a, 8], Theodore D. Satterthwaite [i], Yong Fan [a], Lenore J. Launer [c], Colin L. Masters [d], Paul Maruff [d], Chuanjun Zhuo [c, f], Henry Völzke [h], Sterling C. Johnson [i], Jurgen Fripp [j], Nikolaos Koutsouleris [k], Daniel H. Wolf [l], Raquel Gur [8, l], Ruben Gur [8, l], John Morris [m], Marilyn S. Albert [n], Hans J. Grabe [o], Susan M. Resnick [p], R. Nick Bryan [q], David A. Wolk [b], Russell T. Shinohara [a, r, s], Haochang Shou [a, r, 2], Christos Davatzikos [a, 2, 2]

Jinglei Lv, Andrew Zalesky, Lianne Schmaal, Andre Marquand

doi: https://doi.org/10.1101/2021.02.09.430363

This article is a preprint and has not been certified by peer review [what does this mean?].

ONE

rning approach to facilitate ComBat-base
n of multicentre radiomic features in ne

Lucia, Ingrid Masson, Ronan Abgral, Joanne Alfieri, Caroline Rousseau, Augustin Men
dier, Ulrike Schick, Dimitris Visvikis, Mathieu Hatt

NeuroImage
Volume 245, 15 December 2021, 118703

ELSEVIER

A multi-scanner neuroimaging data harmonization using RAVEL and ComBat

Mahbaneh Eshaghzadeh Torbati [a], Davneet S. Minhas [b], Ghasan Ahmad [c], Erin E. O'Connor [c], John Muschelli [d], Charles M. Laymon [b], Zixi Yang [b], Ann D. Cohen [e], Howard J. Aizenstein [e], William E. Klunk [e], Bradley T. Christian [f], Seong Jae Hwang [a, 8], Ciprian M. Crainiceanu [d], Dana L. Tudorascu [e, h]

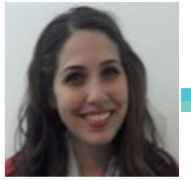
Harmonization Needs Care - Confounded Cases

ENIGMA HIV

Table 1. Demographic and Clinical Information by Study and Scanning Site

Site	Total, No.	Male, No. (%)	Age, mean (SD) [range], y	No. (%) of participants: Taking cART	No. (%) of participants: Detectable viral load
HIVNC (7 sites), US	218	187 (86)	48.5 (8.3) [24-71]	218 (100)	61 (28) [n = 217]
Site 1: University of California, San Diego	21	21 (100)	47.6 (5.0) [37-56]	21 (100)	13 (65) [n = 20]
Site 2: Harbor UCLA Medical Center, Los Angeles, California	52	42 (81)	46.6 (9.1) [24-70]	52 (100)	8 (15)
Site 3: Stanford University, Stanford, California	10	9 (90)	46.4 (10.4) [31-62]	10 (100)	0
Site 4: Colorado	35	34 (97)	49.2 (7.9) [31-66]	35 (100)	14 (40)
Site 5: Pittsburgh, Pennsylvania	19	18 (95)	49.9 (9.7) [34-71]	19 (100)	4 (21)
Site 6: Rochester University, Rochester, New York	41	26 (63)	48.4 (7.7) [26-62]	41 (100)	11 (27)
Site 7: University of California, Los Angeles	40	37 (93)	50.7 (8.4) [28-69]	40 (100)	11 (28)
University of Hawaii, Honolulu, US[17]	175	159 (91)	47.6 (10.5) [20-74]	157 (90)	54 (33) [n = 164]
University of Hawaii, Honolulu, US[18]	53	45 (85)	50.9 (8.0) [40-71]	53 (100)	6 (11)
University of California, San Francisco, US	50	49 (98)	63.6 (2.5) [60-69]	49 (98)	14 (29) [n = 49]
Brown University, Providence, Rhode Island, US	79	48 (61)	45.2 (9.5) [23-65]	66 (84)	22 (28)
University of California, Los Angeles, US[19]	12	12 (100)	46.2 (8.5) [26-57]	9 (75)	6 (50)
University of California, Los Angeles, US[20]	51	46 (90)	50.9 (13.2) [24-76]	51 (100)	21 (41)
University of New South Wales, New South Wales, Australia[21]	39	38 (97)	52.8 (8.2) [39-75]	39 (100)	10 (26)
University of New South Wales, New South Wales, Australia[22]	68	68 (100)	55.3 (6.7) [44-69]	68 (100)	1 (1)
SEARCH-011 Consortium, Bangkok, Thailand	61	26 (43)	34.2 (7.0) [22-56]	0	61 (100)
University of Cape Town, South Africa	181	25 (14)	32.4 (5.0) [22-46]	0	148 (97) [n = 152]
Nice University, Nice, France	155	122 (79)	45.4 (10.0) [22-81]	126 (76)	56 (36)
University of Novi Sad, Novi Sad, Serbia	61	55 (90)	44.3 (10.9) [25-66]	61 (100)	10 (16)

Abbreviations: cART, combination antiretroviral therapy; HIVNC, HIV Neuroimaging Consortium; SEARCH, South East Asian Rese



Talia Nir, PhD



JAMA Network Open

Nir et al / Jahanshad 2021

Original Investigation | Infectious Diseases

Association of Immunosuppression and Viral Load With Subcortical Brain Volume in an International Sample of People Living With HIV

Talia M. Nir, PhD; Jean-Paul Fouche, PhD; Jintanat Ananworanich, MD, PhD; Beau M. Ances, MD, PhD; Jasmina Boban, MD, PhD; Bruce J. Brew, MD; Joga R. Chaganti, MD; Linda Chang, MD, MS; Christopher R. K. Ching, PhD; Lucette A. Cysique, PhD; Thomas Ernst, PhD; Joshua Faskowitz, BA; Vikash Gupta, PhD; Jaroslaw Harezlak, PhD; Jodi M. Heaps-Woodruff, PhD; Charles H. Hinkin, PhD; Jacqueline Hoare, PhD; John A. Joska, PhD; Kalpana J. Kallianpur, PhD; Taylor Kuhn, PhD; Hei Y. Lam, BA; Meng Law, MD; Christine Lebrun-Frénay, MD, PhD; Andrew J. Levine, PhD; Lydiane Mondot, MD; Beau K. Nakamoto, MD; Bradford A. Navia, MD, PhD; Xavier Pennec, PhD; Eric C. Porges, PhD; Lauren E. Salminen, PhD; Cecilia M. Shikuma, MD; Wesley Surento, MS; April D. Thames, PhD; Victor Valcour, MD, PhD; Matteo Vassallo, MD; Adam J. Woods, PhD; Paul M. Thompson, PhD; Ronald A. Cohen, PhD; Robert Paul, PhD; Dan J. Stein, PhD; Neda Jahanshad, PhD

A Subcortical volumes and CD4+ cell counts

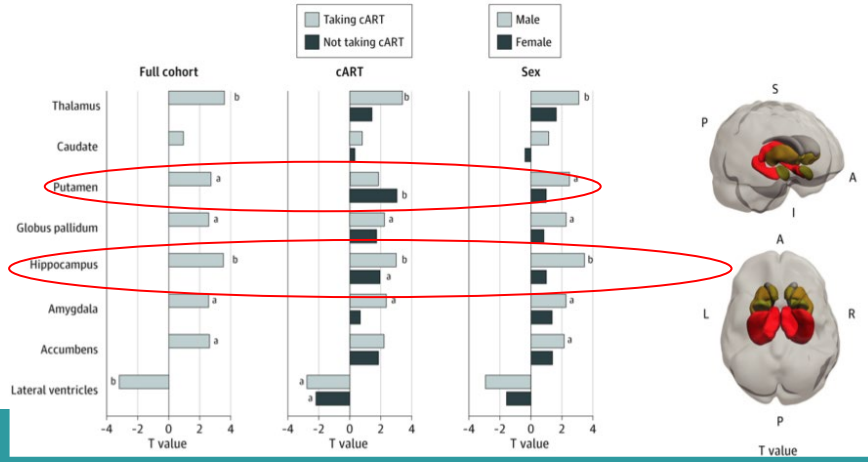
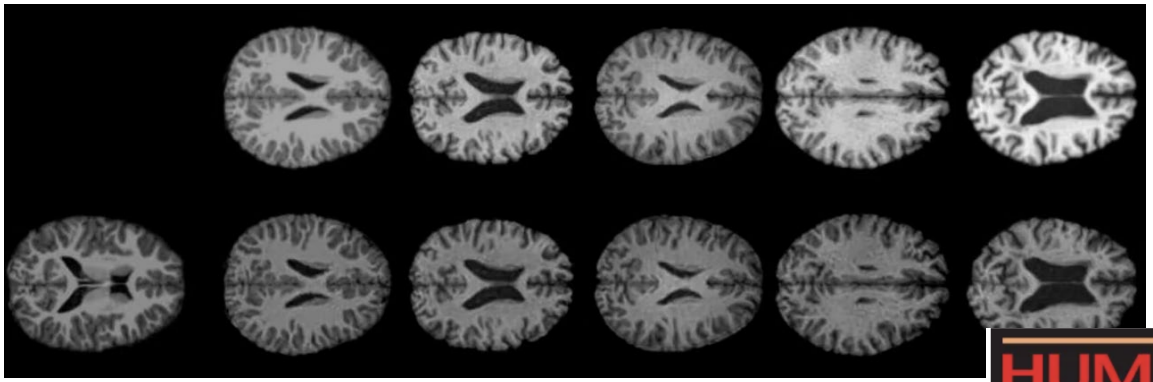


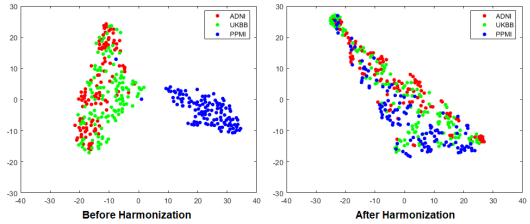
Image Harmonization with AI

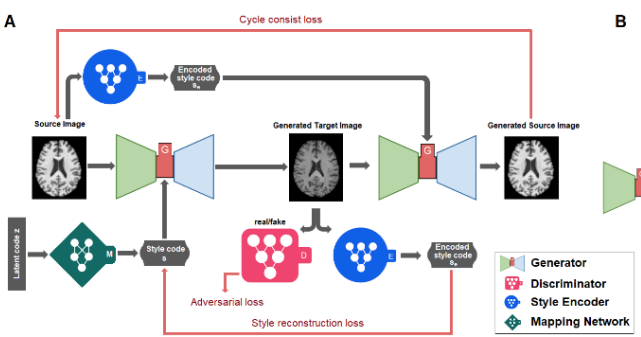


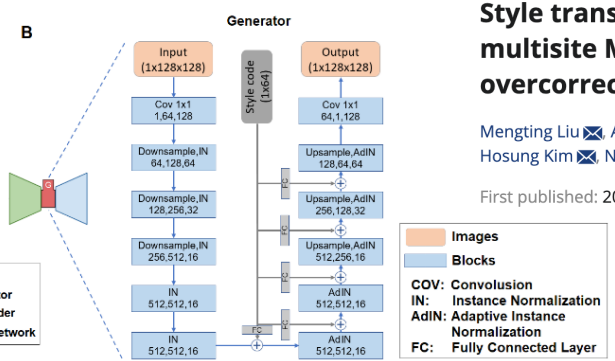
Mengting Liu, PhD



Liu et al HBM 2023







HUMAN BRAIN MAPPING
Open Access

RESEARCH ARTICLE | Open Access

Style transfer generative adversarial networks to harmonize multisite MRI to a single reference image to avoid overcorrection

Mengting Liu, Alyssa H. Zhu, Piyush Maiti, Sophia I. Thomopoulos, Shruti Gadewar, Yaqiong Chai, Hosung Kim, Neda Jahanshad, for the Alzheimer's Disease Neuroimaging Initiative

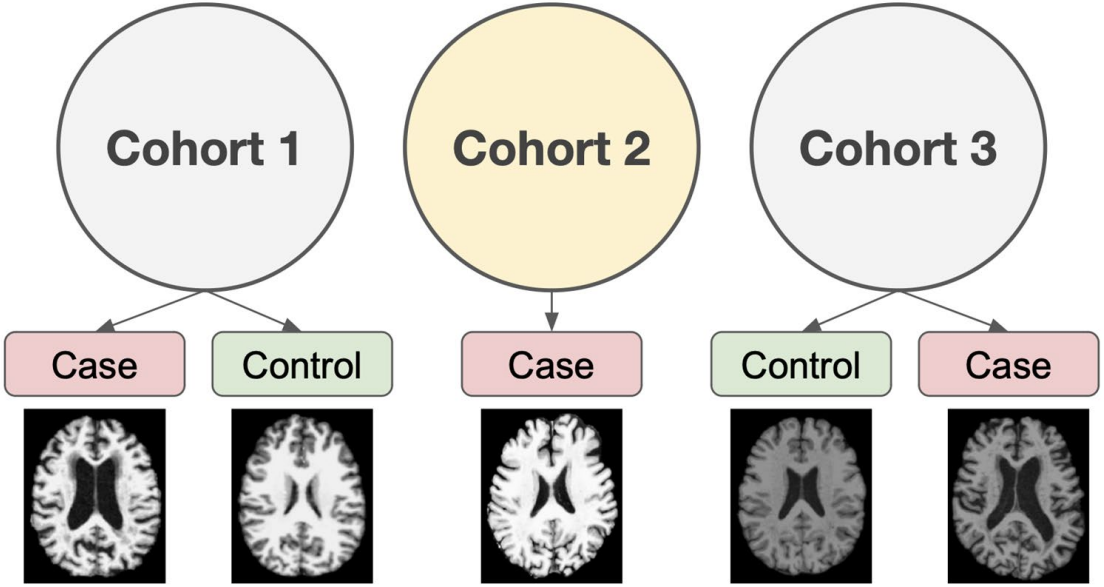
First published: 20 July 2023 | https://doi.org/10.1002/hbm.26422

Site effects how-to and when: An overview of retrospective techniques to accommodate site effects in multi-site neuroimaging analyses

Johanna M M Bayer [1,2], Paul M Thompson [3], Christopher R K Ching [3], Mengting Liu [4], Andrew Chen [5,6], Alana C Panzenhagen [7,8], Neda Jahanshad [9], Andre Marquand [10], Lianne Schmaal [1,2], Philipp G Sämann [11]

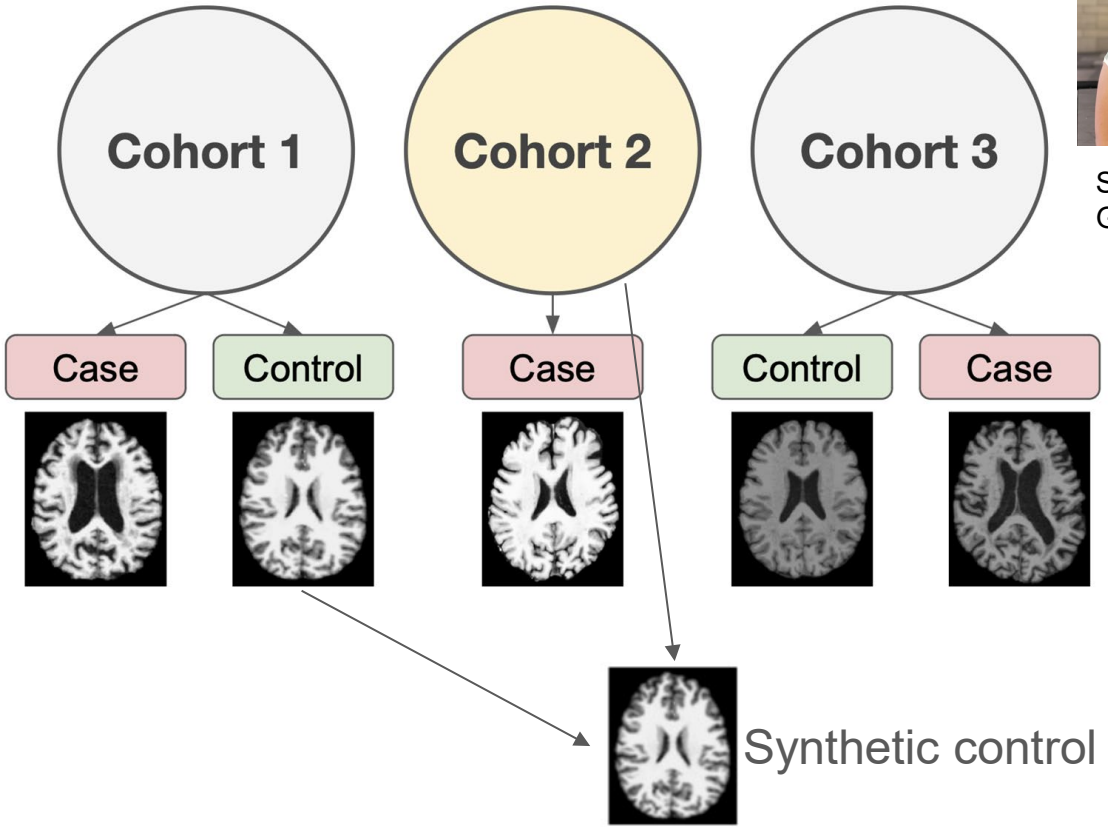
Using generated MR scans

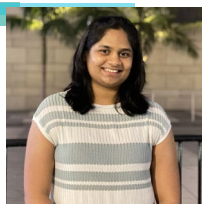
	Control	Cases
Study 1		
Study 2		
Study 3		
Study 4		



Generative AI to Create Synthetic Controls

	Control	Cases
Study 1		
Study 2		
Study 3		
Study 4		



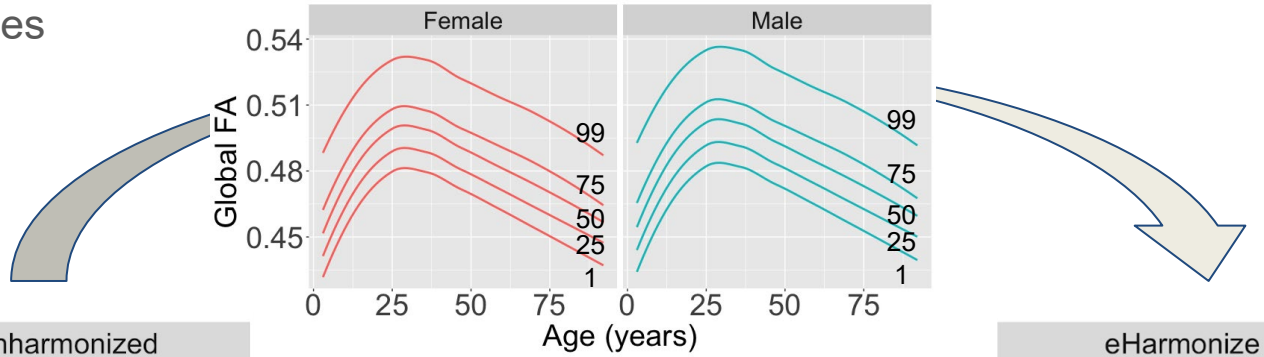


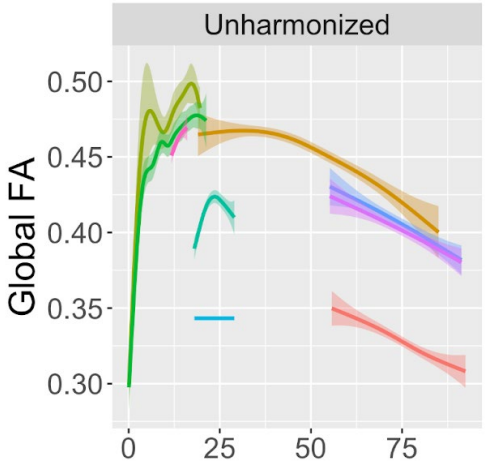
Shruti Gadewar

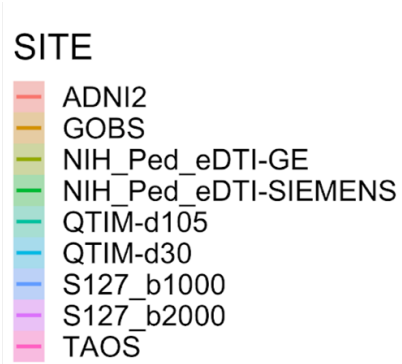
Generative AI to Create Synthetic Controls

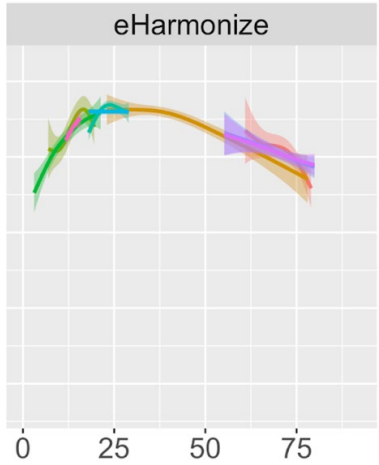
Use statistical harmonization methods on generated images to determine scale and shift

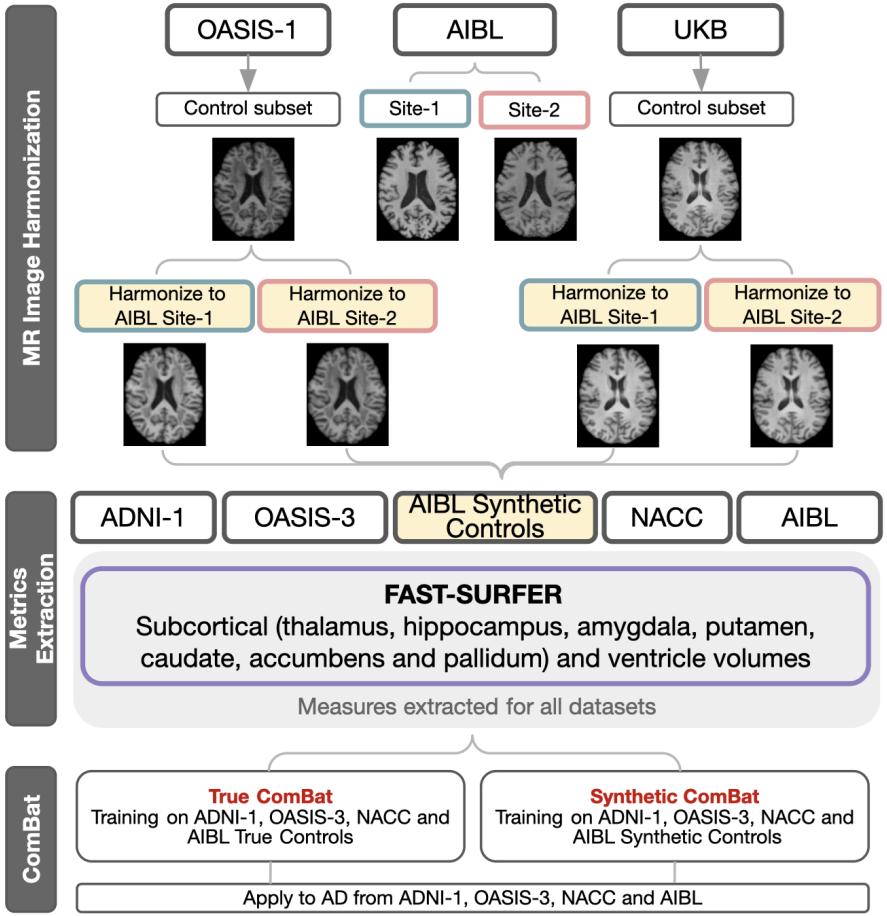
And apply transformation to cases

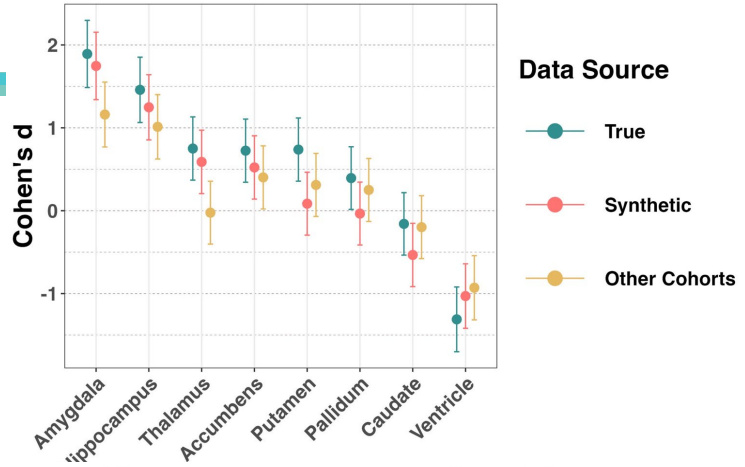


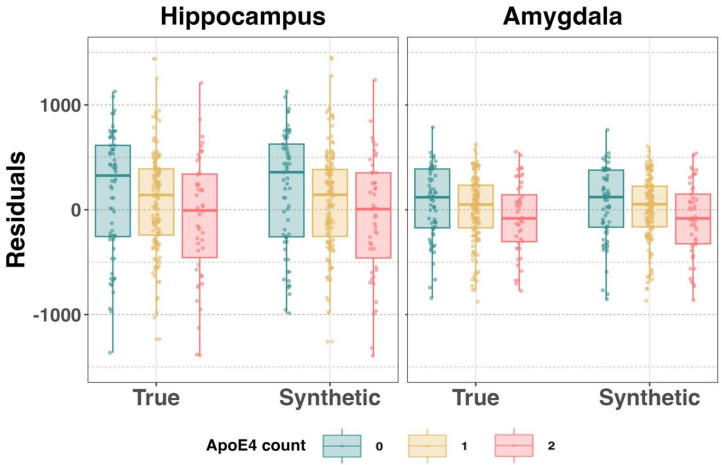







MR Image Harmonization
OASIS-1
AIBL
UKB
Control subset
Site-1
Site-2
Control subset
Harmonize to AIBL Site-1
Harmonize to AIBL Site-2
Harmonize to AIBL Site-1
Harmonize to AIBL Site-2
Metrics Extraction
ADNI-1
OASIS-3
AIBL Synthetic Controls
NACC
AIBL
FAST-SURFER
Subcortical (thalamus, hippocampus, amygdala, putamen, caudate, accumbens and pallidum) and ventricle volumes
Measures extracted for all datasets
ComBat
True ComBat
Training on ADNI-1, OASIS-3, NACC and AIBL True Controls
Synthetic ComBat
Training on ADNI-1, OASIS-3, NACC and AIBL Synthetic Controls
Apply to AD from ADNI-1, OASIS-3, NACC and AIBL

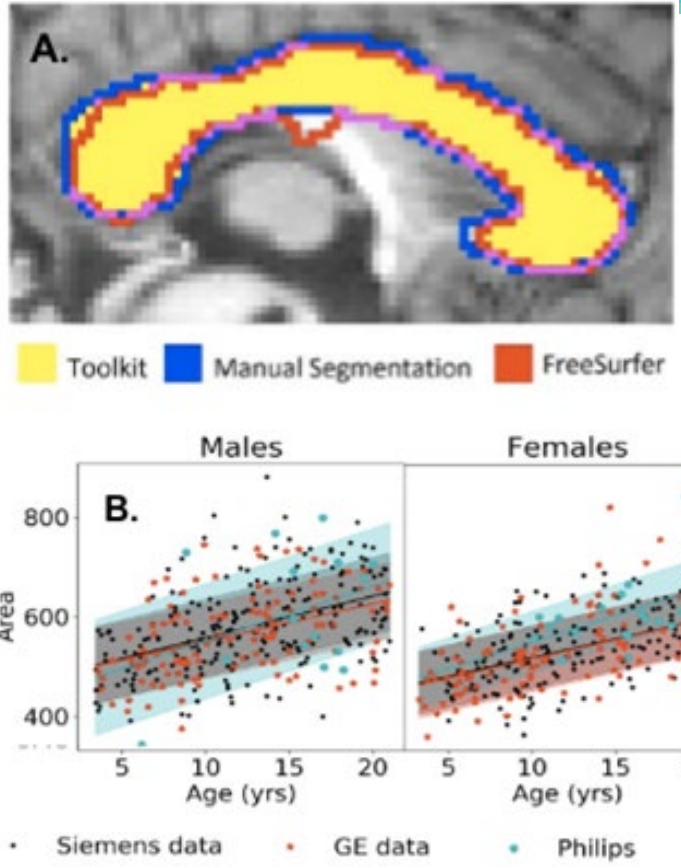

Cohen's d
2
1
0
-1
Amygdala
Hippocampus
Thalamus
Accumbens
Putamen
Pallidum
Caudate
Ventricle
Data Source
True
Synthetic
Other Cohorts

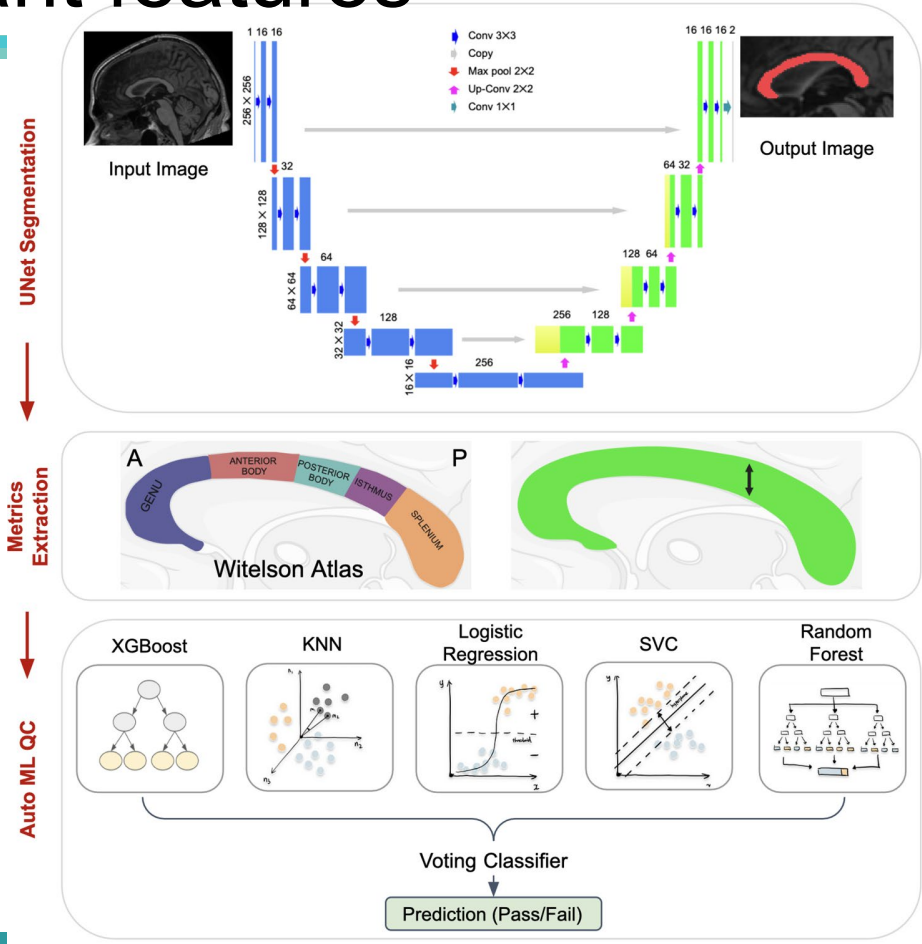

Hippocampus
Amygdala
Residuals
1000
0
-1000
True
Synthetic
True
Synthetic
ApoE4 count
0
1
2

Can we minimize the need for harmonization altogether

?

Towards scanner invariant features

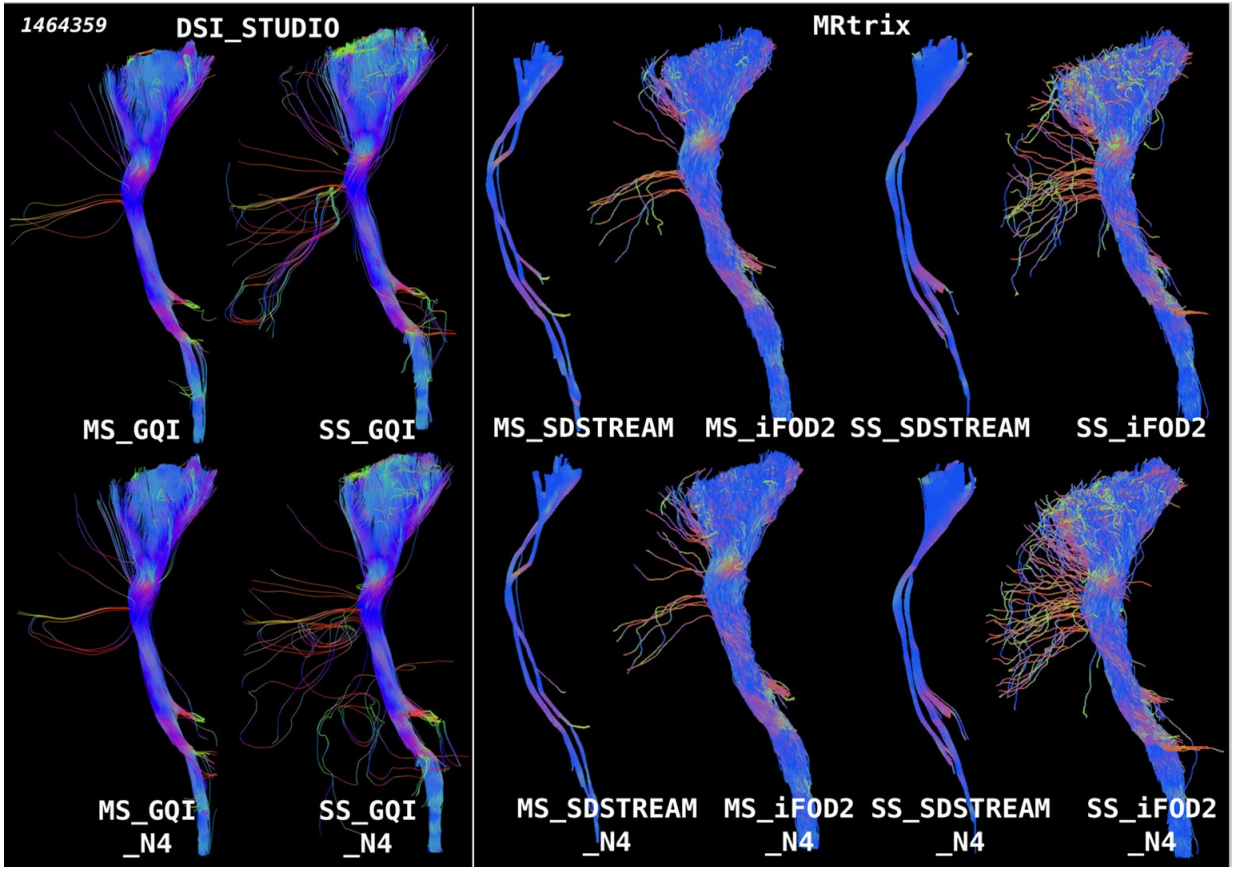




Zhu et al SIPAIM 2018

Bhatt, Gadewar et al Nat Comm 2025

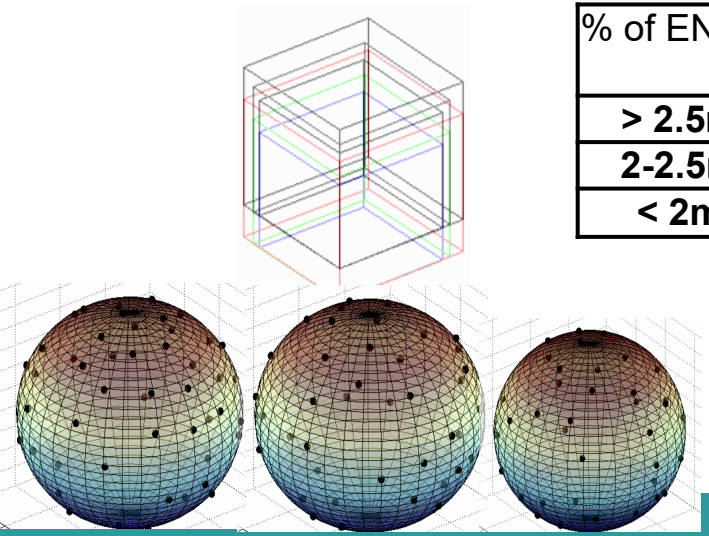
Can we standardize tractography based connectivity analyses?



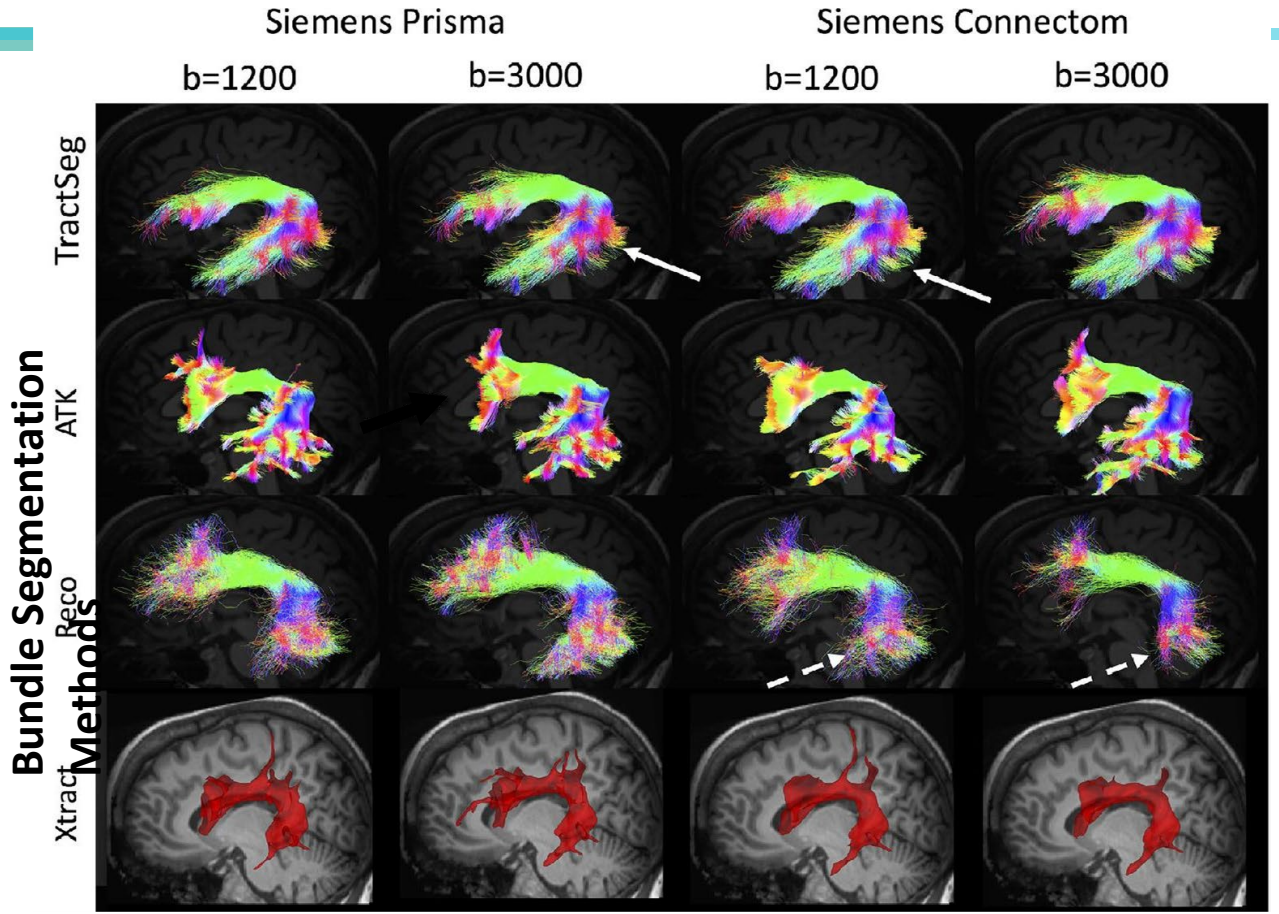
Can we standardize tractography based connectivity analyses?

Acquisition protocol diversity

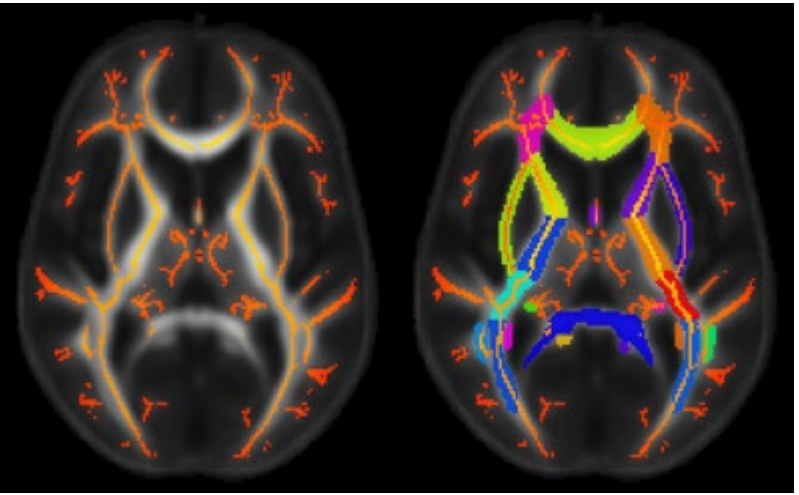
Much of the clinically focused datasets are optimized to collect information from multiple domains and may limit time in the scanner, so acquisition protocols are not always optimal for all analyses



% of ENIGMA	**DTI** (<50) b<=2000	**HARDI** b<=2000	**Multishell** b<=2000	**DTI** b>=3000	**HARDI** b>=3000	**Multishell** b>=3000
> 2.5mm	~10%	~2%				
2-2.5mm	**~70%**	~10%	~5%	~2%		
< 2mm						~1%



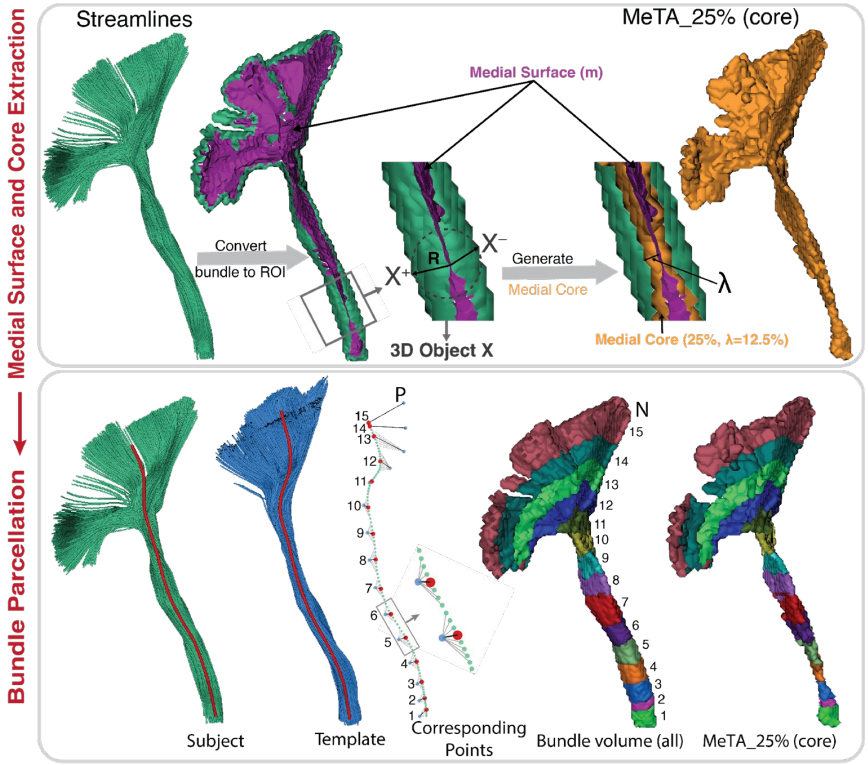
Schilling et al. 2021





Medial Tractography Analysis (MeTA)

- Reduce microstructural heterogeneity and variability of dMRI metrics within white matter bundles
- Preserve bundle shape while capturing regional variation within and along white matter bundles
- Establish regional correspondence for along-tract comparisons across subjects and acquisition protocols



Ba Gari et al, 2023, 2024, in progress

Volumetric “core” analysis

Converted streamlines of WM bundle into a binary mask.

Generate medial surface *m* by using the concept of a Maximum Inscribed Ball at each point on *m* of bundle shape (Yushkevich et al. 2006)

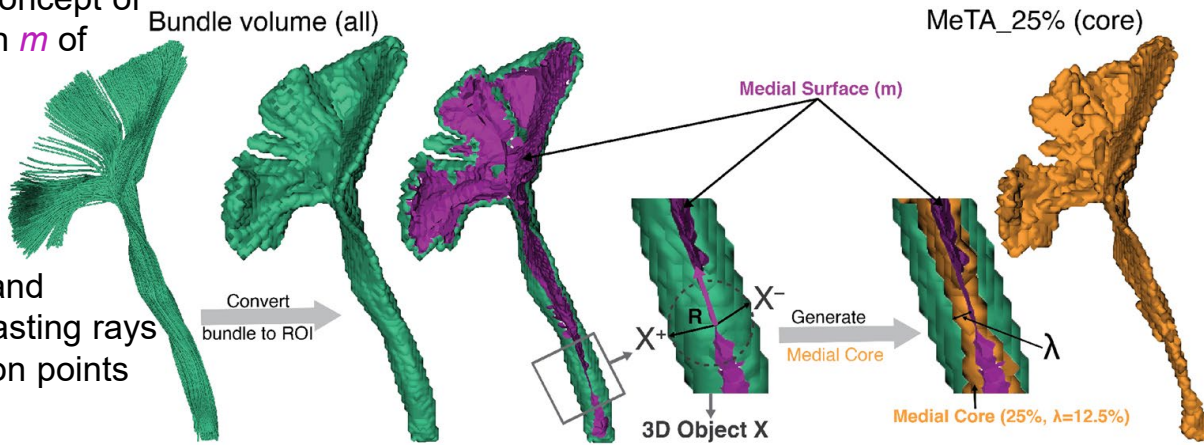
$$X^{\pm}(u) = m(u) + R(u)U^{\pm}(u)$$

$$U^{\pm} = -\nabla_m R \pm \sqrt{1 - \| \nabla_m R \|^2}\, N_m$$

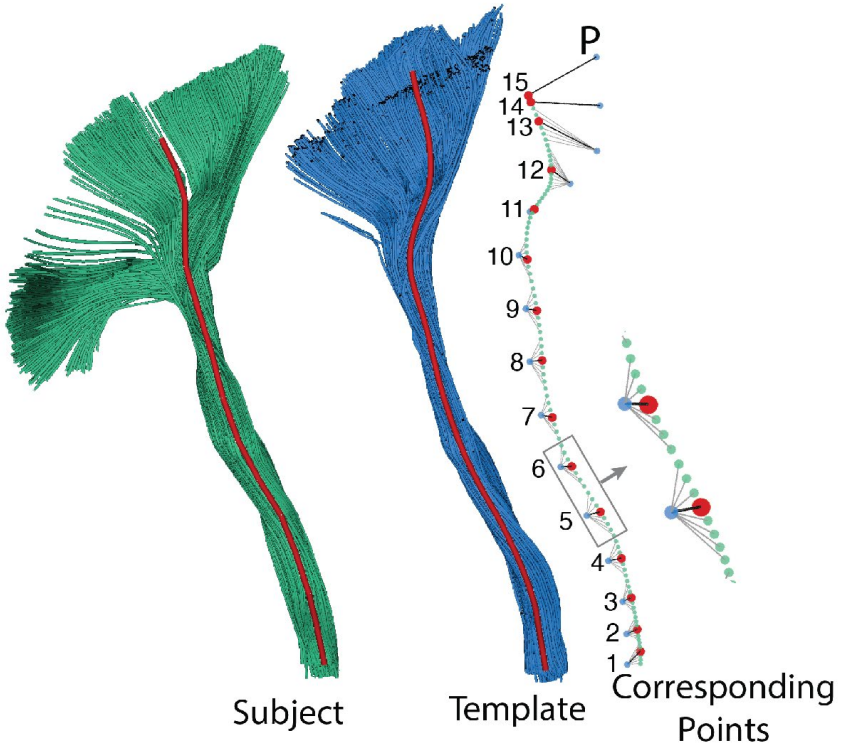
Calculated normal at vertex points on *m* and identified corresponding points on $X^{\pm}$ by casting rays along the normal vectors to find intersection points on $X^{\pm}$ (Sullivan et al., 2019)

$$P^{\pm}(u) = m(u) + \lambda N(u)$$

Voxelized MeTA_25% (core) using the computed depth from the medial surface



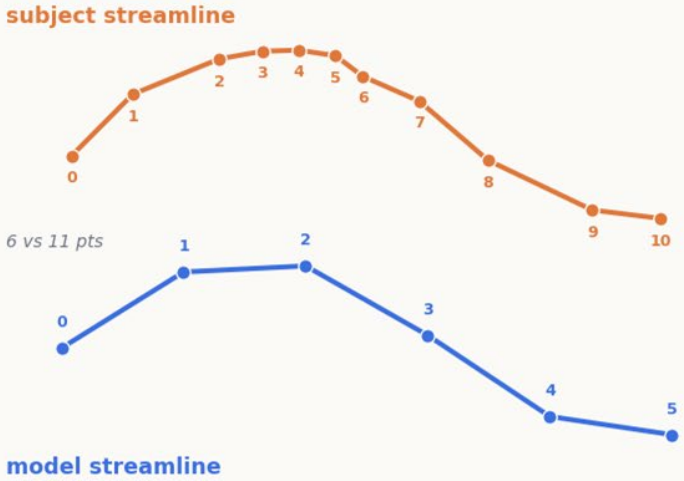
Cross-subject bundle correspondence

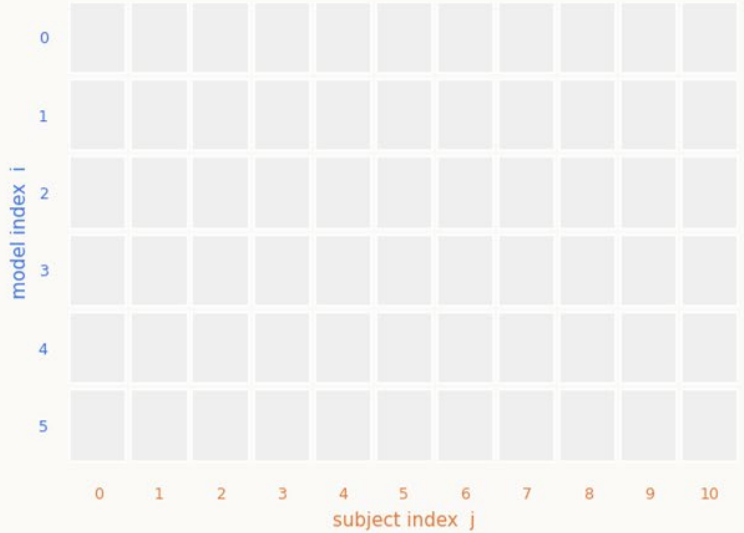


- Generated a centroid with P points along the template bundle length for reference in cross-subject comparisons
- Generated a centroid for each subject's bundle using QuickBundles clustering (Garyfallidis et al., 2012)
- Applied Dynamic Time Warping (DTW) for aligning points between subject and template centroids (Sakoe et al., 1978)

Dynamic Time Warping

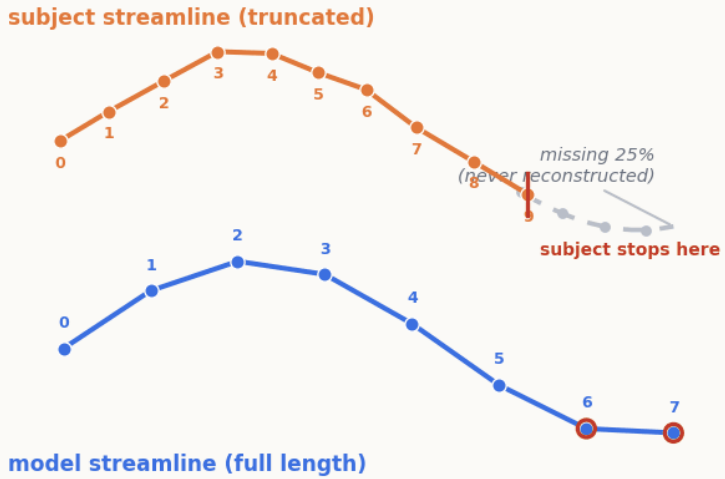
meta/core/corresponds.py · tslearn.metrics.dtw_path

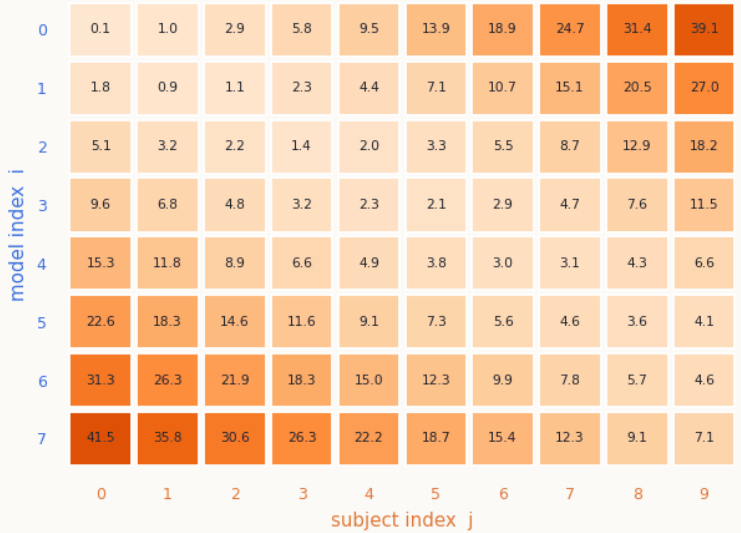




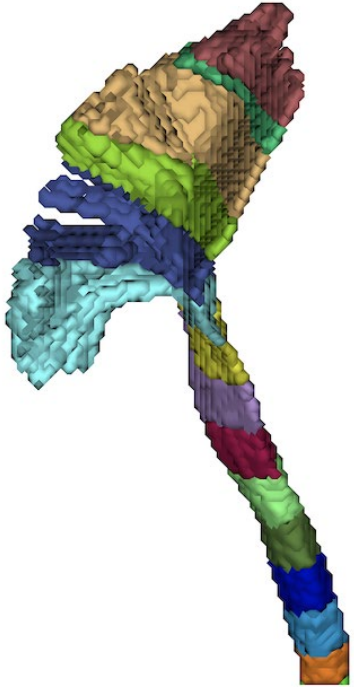
Dynamic Time Warping

meta/core/corresponds.py · tslearn.metrics.dtw_path



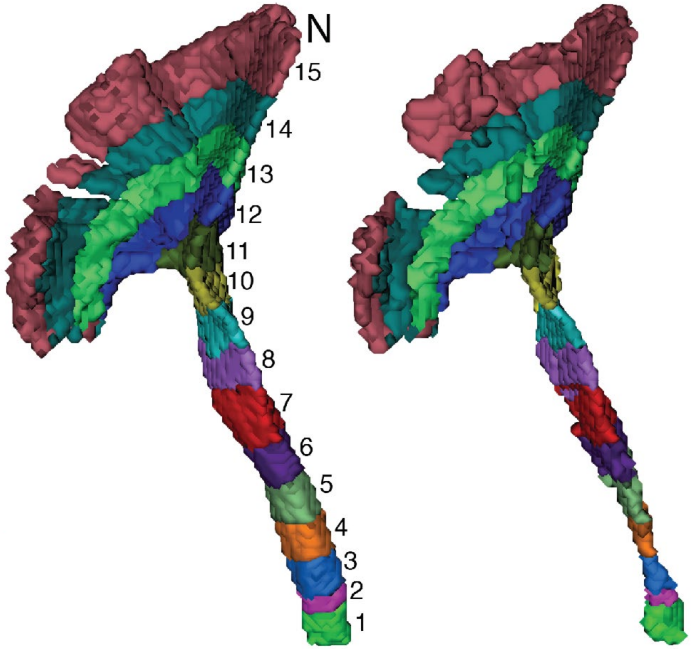


Along-tract analysis



$$Number\ of\ segments = \frac{Average\ bundle\ length\ (mm)}{voxel\ size\ (mm)\ x\ required\ thikness}$$

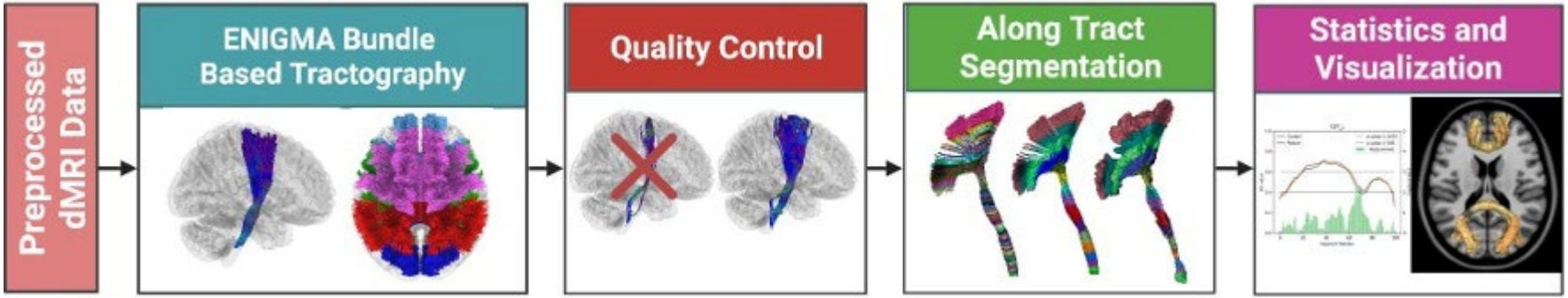
Along-tract analysis



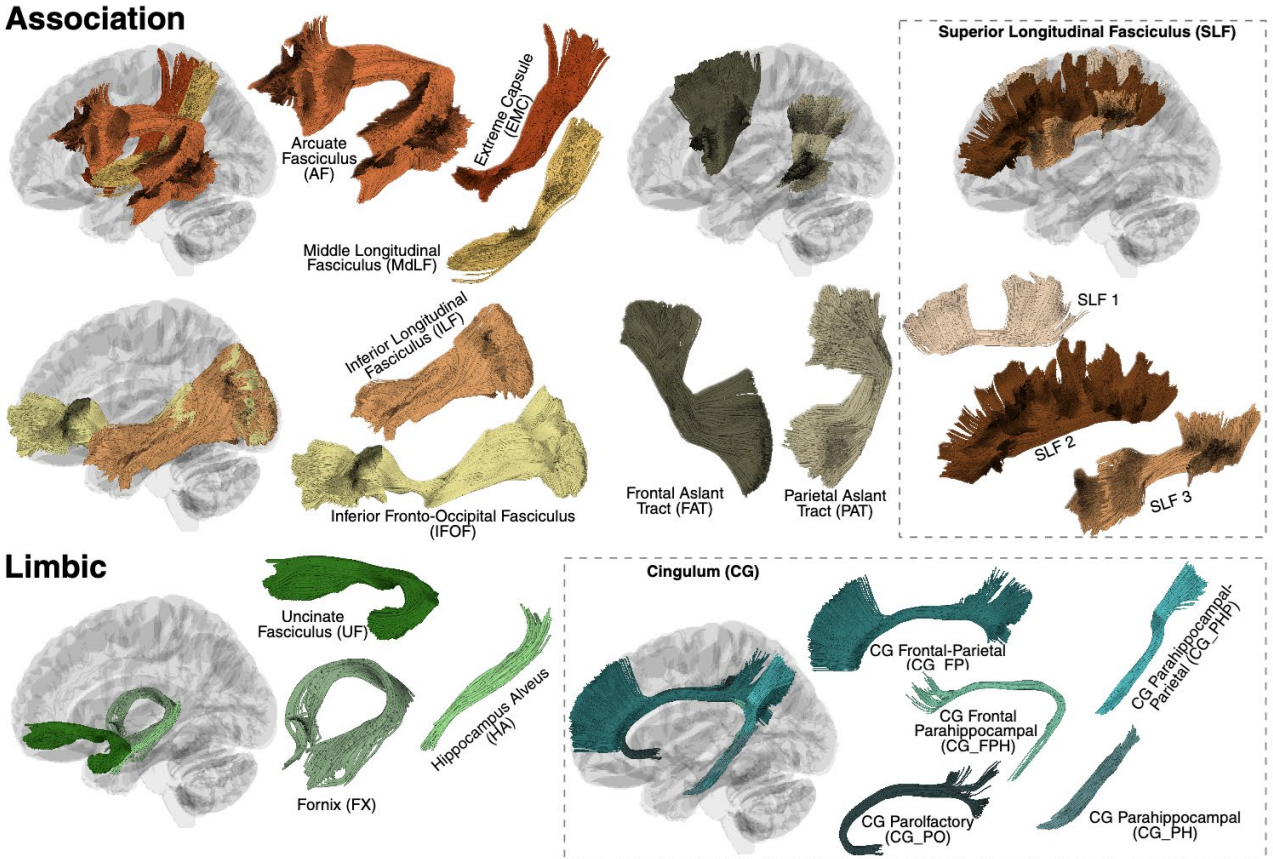
Bundle volume (all) MeTA_25% (core)

ENIGMA Tractometry Toolbox

Standardized processing and quality control



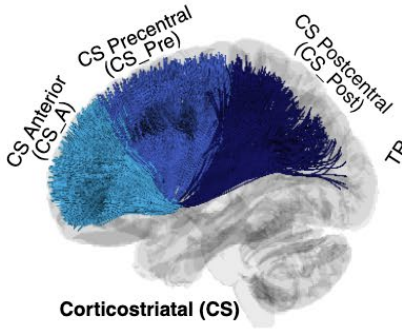
ENIGMA Symmetric Atlas



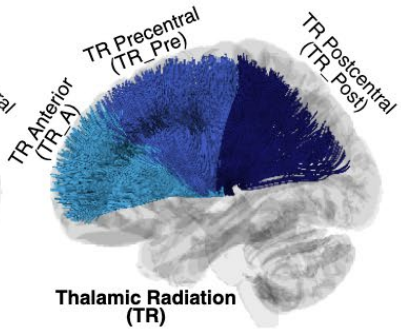
Ba Gari, I., Davis Alibrando, J., Feng, Y., Villalón-Reina, J.E., Cabalinan, A., Nabulsi, L., Thompson, P.M., Yeh, F.-C., Chandio, B.Q., Nir, T.M. and Jahanshad, N., 2026. *ENIGMA Symmetric White Matter Tractography Atlas (Version 1.0.0)* [data set]. Zenodo. Available at: https://zenodo.org/records/19656193

ENIGMA Symmetric Atlas

Projection Basal Ganglia



Corticostriatal (CS)

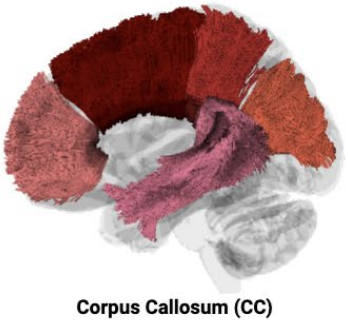


Thalamic Radiation (TR)



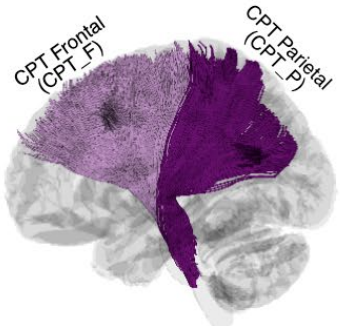
Optic Radiation (OR)

Commissural

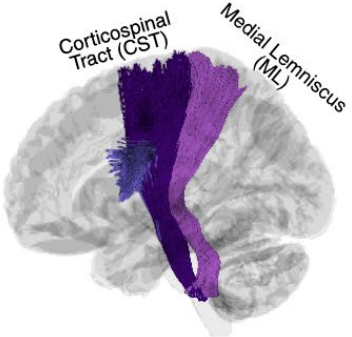


Corpus Callosum (CC)

Projection Brainstem

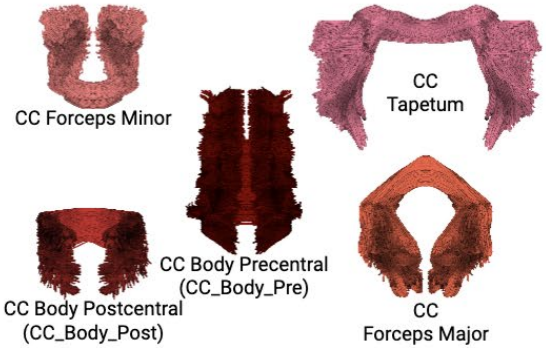


Corticopontine Tract (CPT)



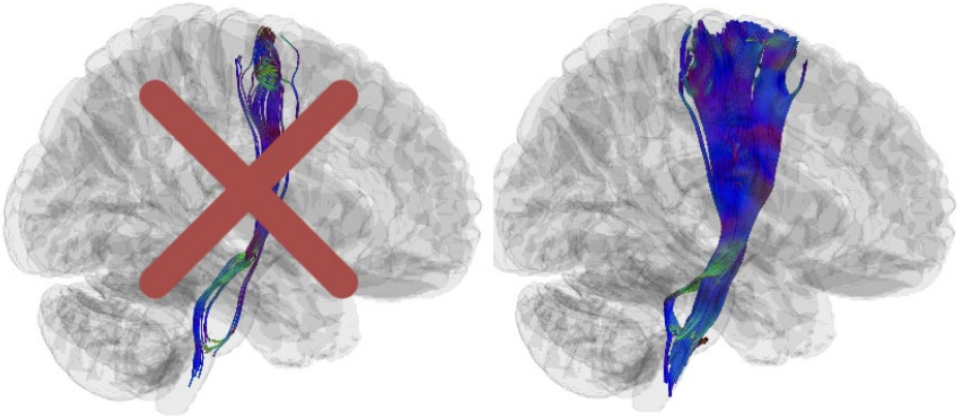


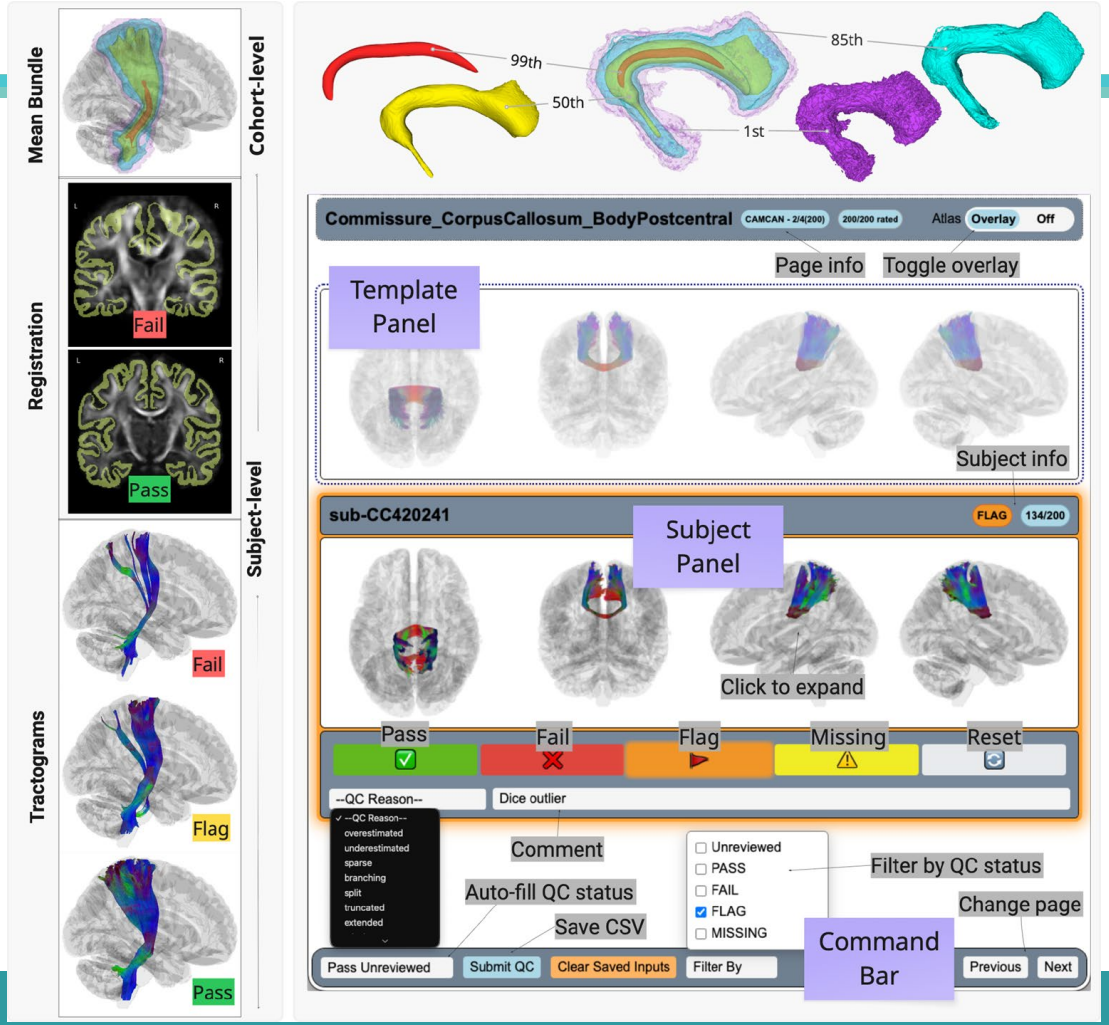
Corticobulbar Tract (CBT)



Ba Gari, I., Davis Alibrando, J., Feng, Y., Villalón-Reina, J.E., Cabalinan, A., Nabulsi, L., Thompson, P.M., Yeh, F.-C., Chandio, B.Q., Nir, T.M. and Jahanshad, N., 2026. *ENIGMA Symmetric White Matter Tractography Atlas (Version 1.0.0)* [data set]. Zenodo. Available at: https://zenodo.org/records/19656193

Quality Control






Talia Nir


Bramsh Chandio


Julio Villalon


Iyad Ba Gari

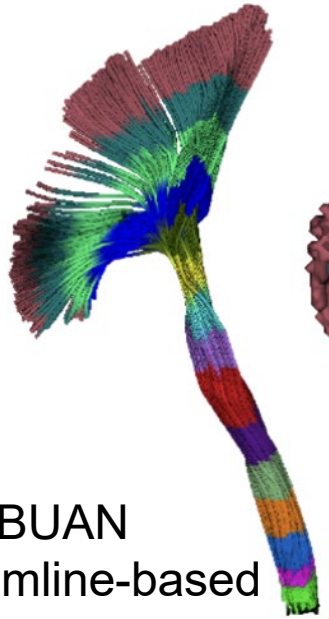



Jonathan Alibrando





Along Tract Segmentation



BUAN
Streamline-based



MeTA
Volumetric

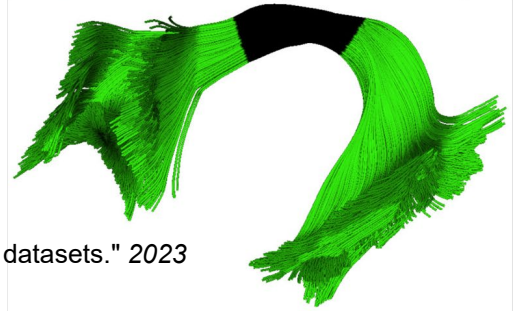




Microstructure Along-Tract Segment

Segment	1	2	3	...	N
Subj 1	0.1	0.3	0.6	...	0.1
Subj 2	0.2	0.3	0.5	...	0.1
...					
Subj N	0.2	0.3	0.7	...	0.2

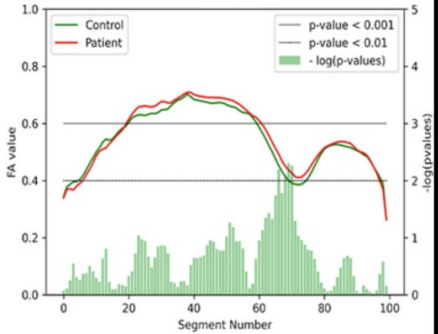
DTI_i ~ TARGET + Age + Sex + (1|Site)

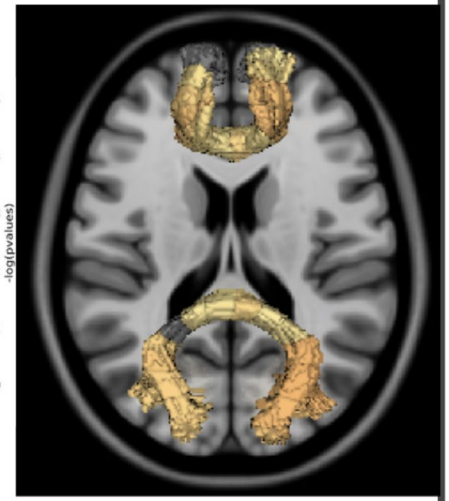


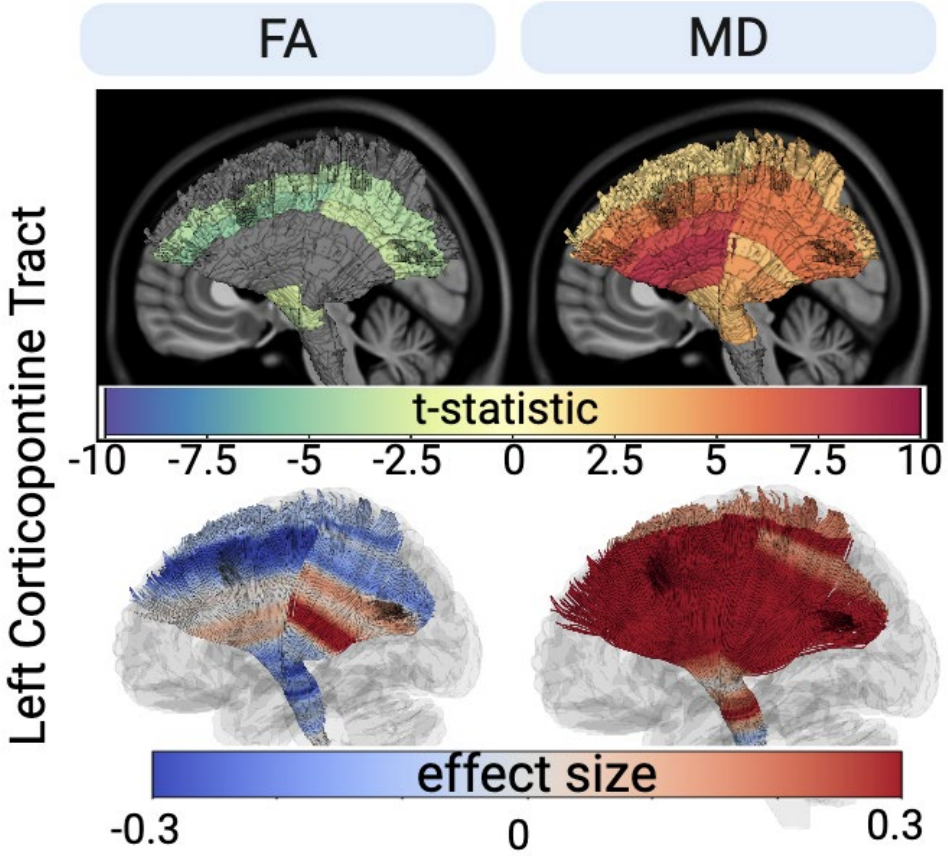
Chandio, B.Q, et al. "Bundle analytics, a computational framework for investigating the shapes and profiles of brain pathways across populations." *Scientific reports* 10.1 (2020): 17149.

Ba Gari, I., et al. "Medial Tractography Analysis (MeTA) for white matter population analyses across datasets." *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2023.

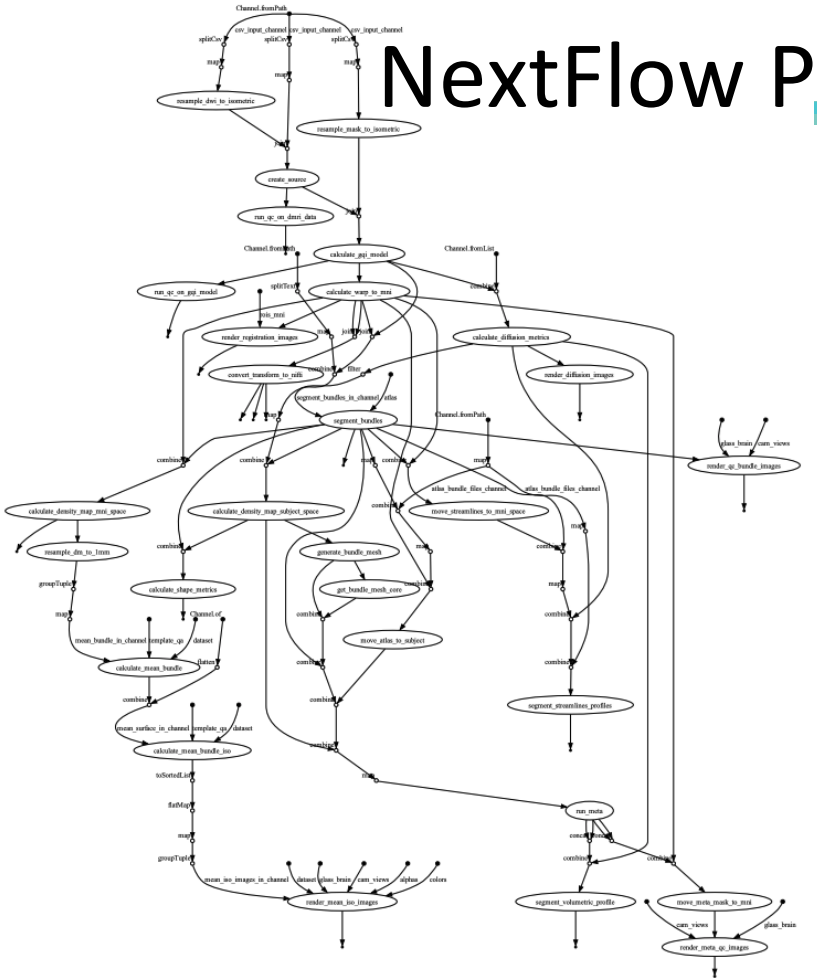
Statistics and Visualization





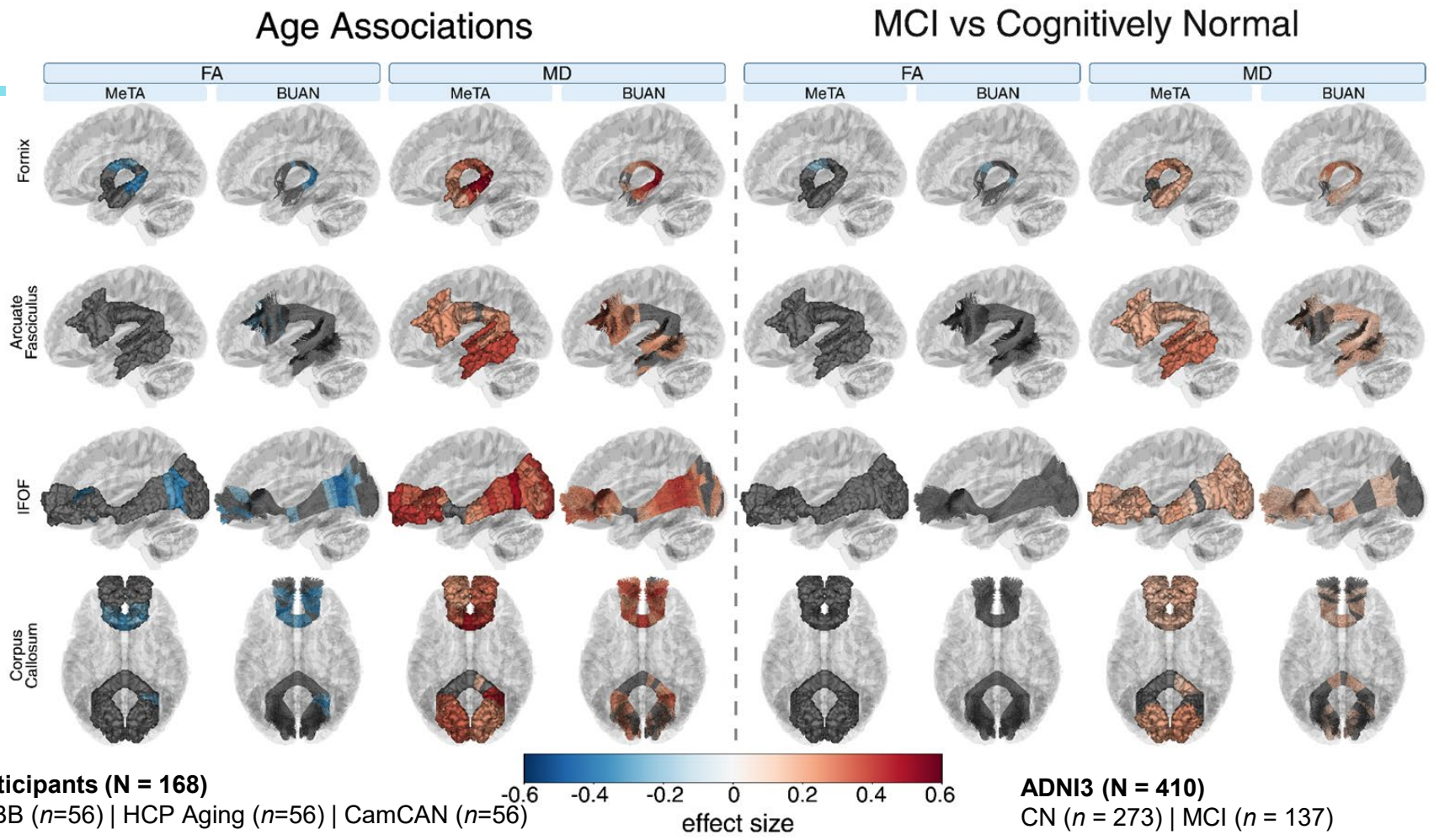


NextFlow Pipeline Usage




```
${project}/tract_pipeline.sh \
    -o "${project}/output" \
    -s "${project}/scripts/subjIDs_HCP_A_small.txt" \
    -i "${project}/data/HCP_A/{subject}/{subject}_dwi.nii.gz" \
    -m "${project}/data/HCP_A/{subject}/{subject}_mask.nii.gz" \
    -c "${project}/data/HCP_A/{subject}/{subject}.bvec" \
    -l "${project}/data/HCP_A/{subject}/{subject}.bval" \
    -d HCP_A -M dti_fa,md \
    -p throttle_light -e sge -q compute9.q -n 300 \
```

https://github.com/ENIGMA-git/eTractometry



Participants (N = 168)
UKBB (*n*=56) | HCP Aging (*n*=56) | CamCAN (*n*=56)

ADNI3 (N = 410)
CN (*n* = 273) | MCI (*n* = 137)

Summary

Multi-study data integration is how we turn local observations into reliable science.

Challenges:
Many axes of data heterogeneity
Data acquisition heterogeneity
Study design heterogeneity
Population heterogeneity

…Data is messy

Opportunities:
Data standardization → invariance
Statistical / AI advances
Population understanding
Uncertainty modeling

Laboratory of Brain eScience (LoBeS)

neda.jahanshad@ini.usc.edu

Supported by:
R01AG087513
R01MH134004
RF1NS136995
R01HD117775
R01NS128486
S10OD032285
Wellcome LEAP CARE
Alz Assoc

Thank you present & former lab members, thousands of ENIGMA participants group leads, investigators & research volunteers!



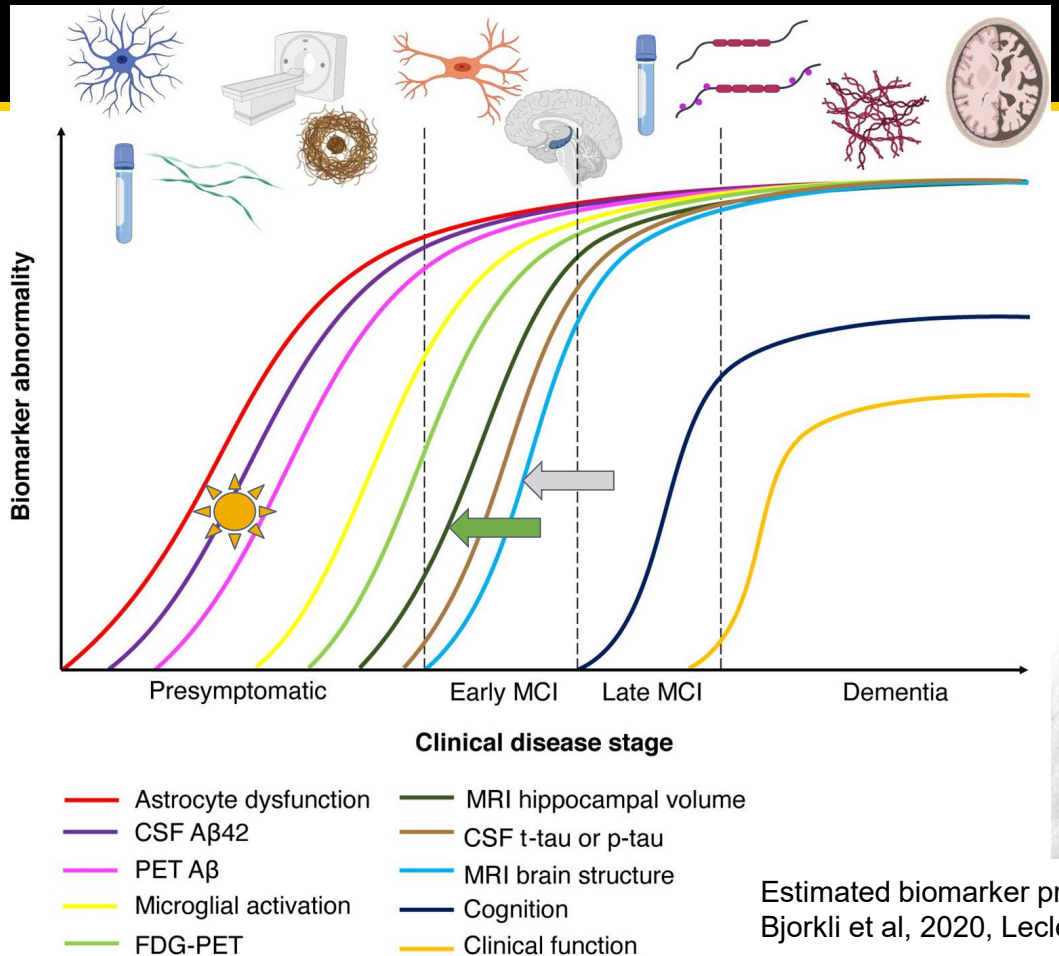




Lab members:

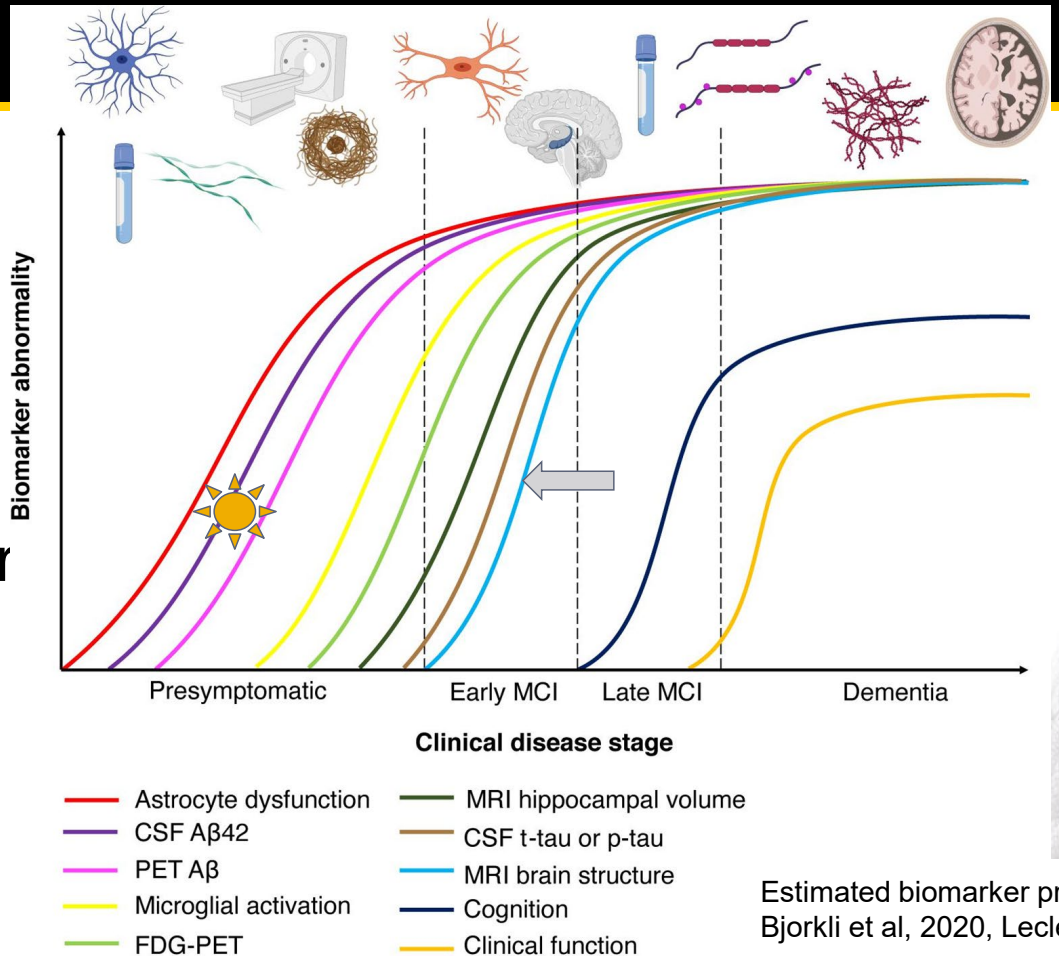
Talia Nir, PhD (Assistant Professor)
Ravi Bhatt, PhD
Elizabeth Haddad, PhD
Saeed Moazami, PhD
Shayan Javid, MS
Iyad Ba Gari, MS
Shruti Gadewar, MS
Kevin Low, MS
JD Alibrando, MS
Sunanda Somu, MS
Aarya Vakharia, MS

Mengting Liu, PhD (former lab member)
Alyssa Zhu, PhD (former lab member)



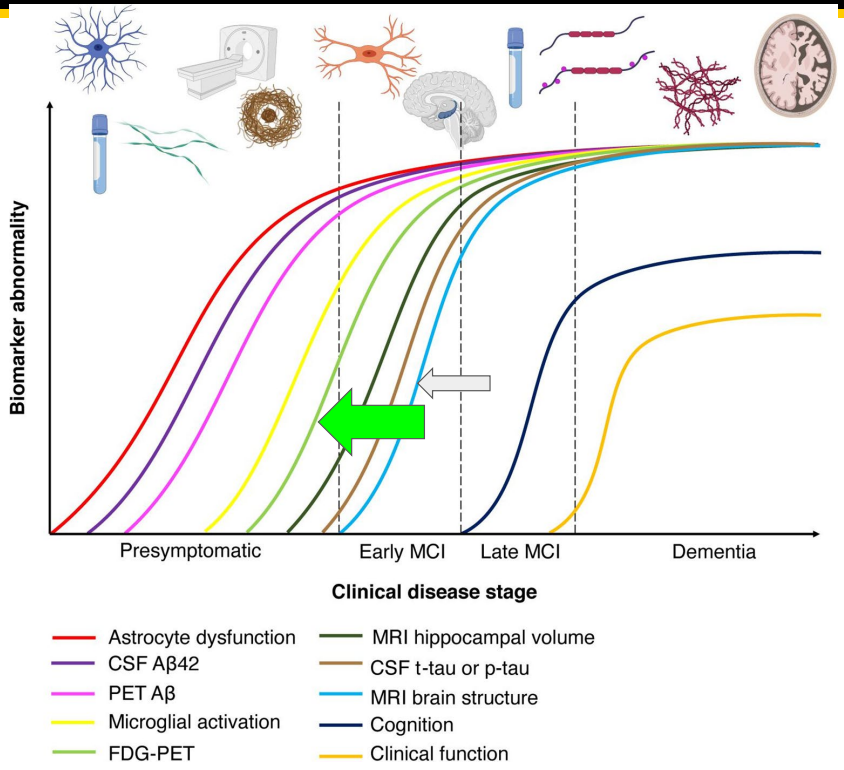
Estimated biomarker progression curves
Bjorkli et al, 2020, Leclerc et al, 2013, Jack et al, 2010

Earlier brain biomarkers for disease

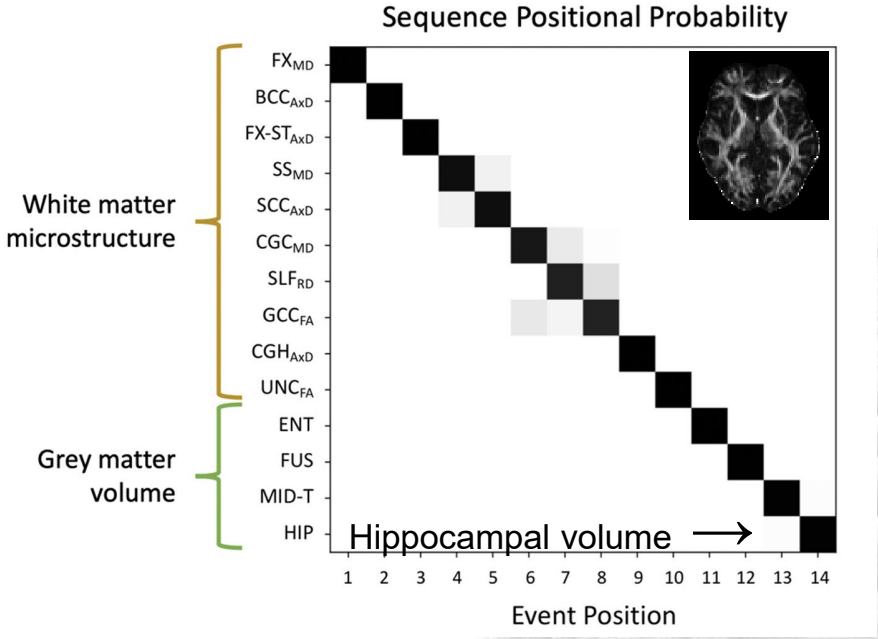


Estimated biomarker progression curves
Bjorkli et al, 2020, Leclerc et al, 2013, Jack et al, 2010

Beyond cortical thickness and hippocampal volume



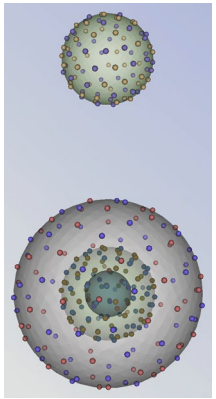
Event based modeling



Parker/Oxtoby et al, 2023 biorxiv

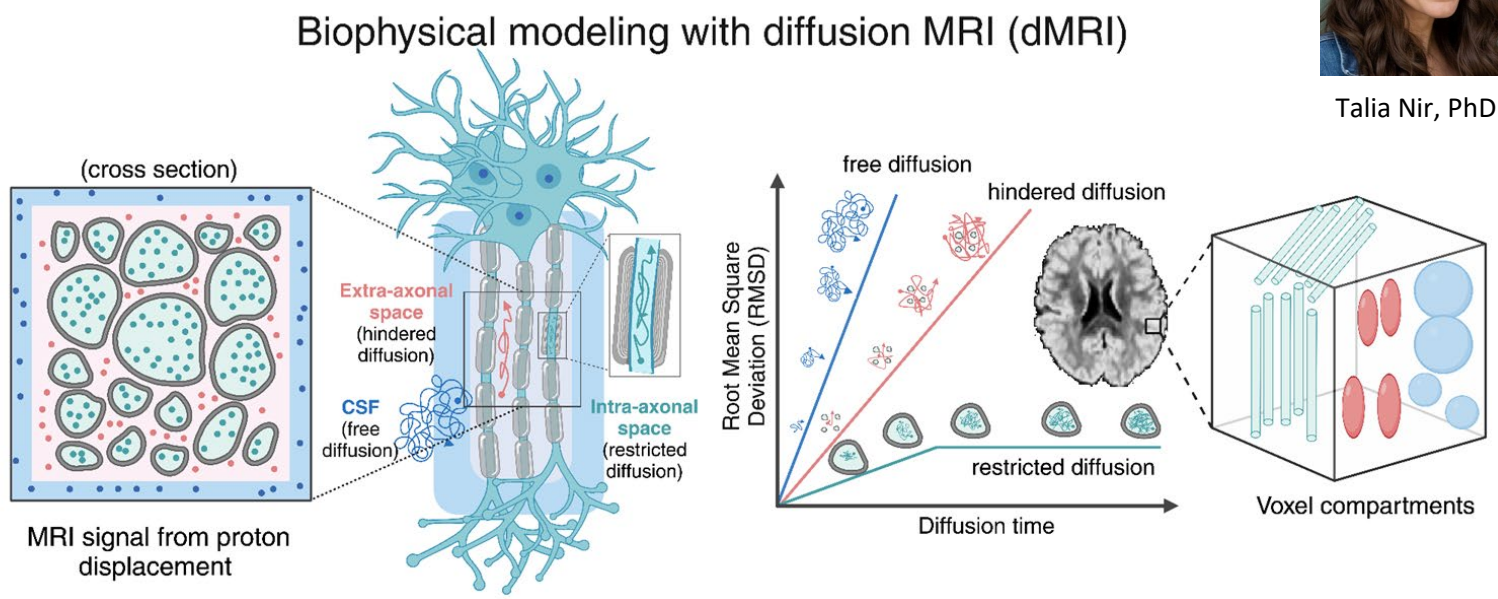
Advanced models of microstructure

"*in-vivo* histology"





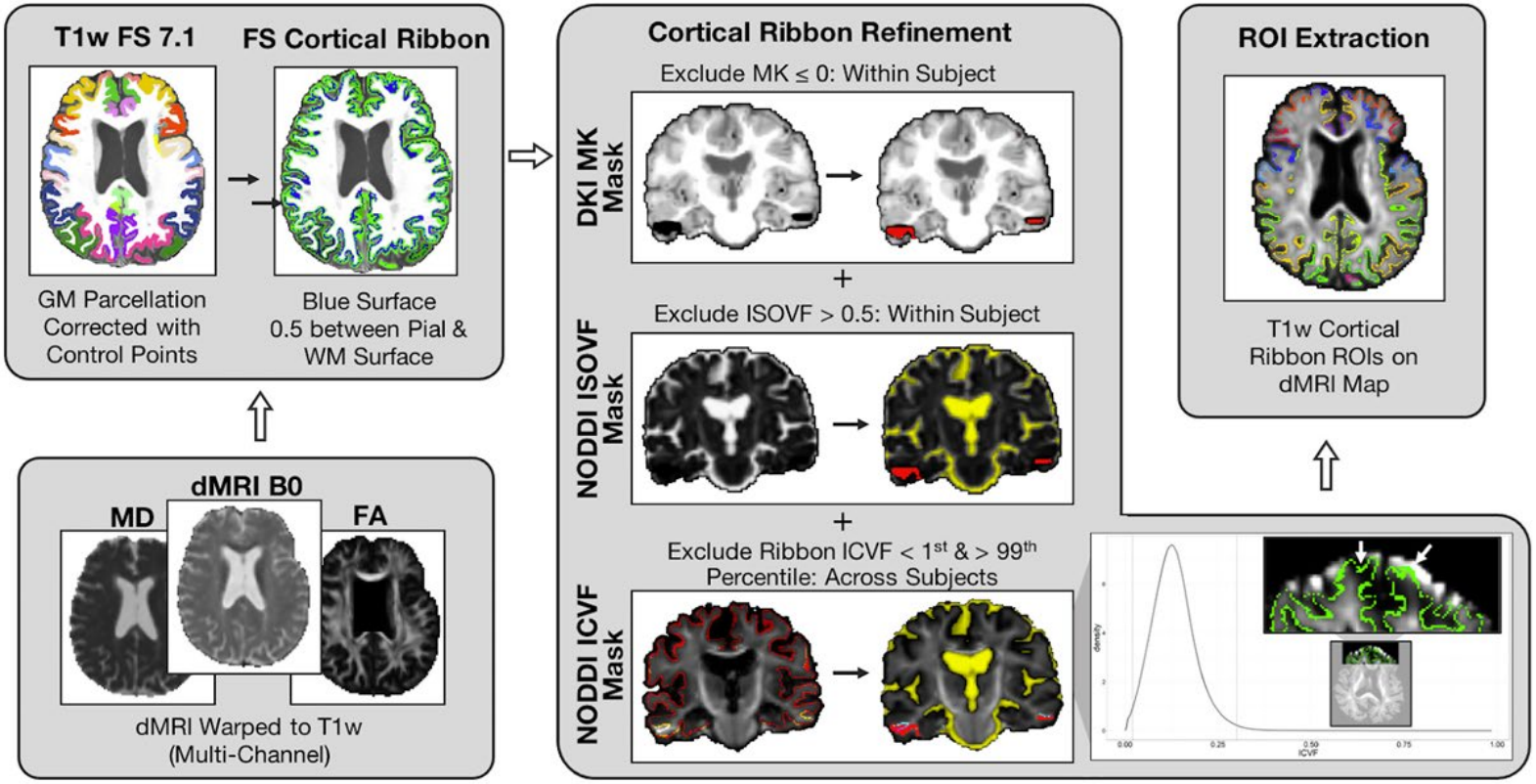
Talia Nir, PhD



Differentiate intracellular from extracellular water diffusion

figure by Elizabeth Haddad, PhD

Microstructure in the gray matter cortex



Can diffusion MRI inform amyloid / tau load?

Molecular Psychiatry

www.nature.com/mp

ARTICLE OPEN

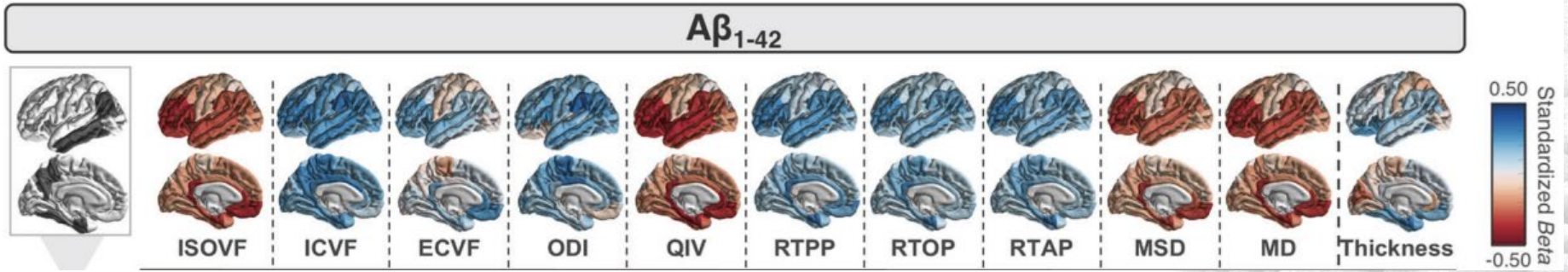
Check for updates

Cortical microstructural associations with CSF amyloid and pTau

Talia M. Nir[1]✉, Julio E. Villalón-Reina[1], Lauren E. Salminen[1], Elizabeth Haddad[1], Hong Zheng[1], Sophia I. Thomopoulos[1], Clifford R. Jack Jr.[2], Michael W. Weiner[3], Paul M. Thompson[1], Neda Jahanshad[1] for the Alzheimer's Disease Neuroimaging Initiative (ADNI)*

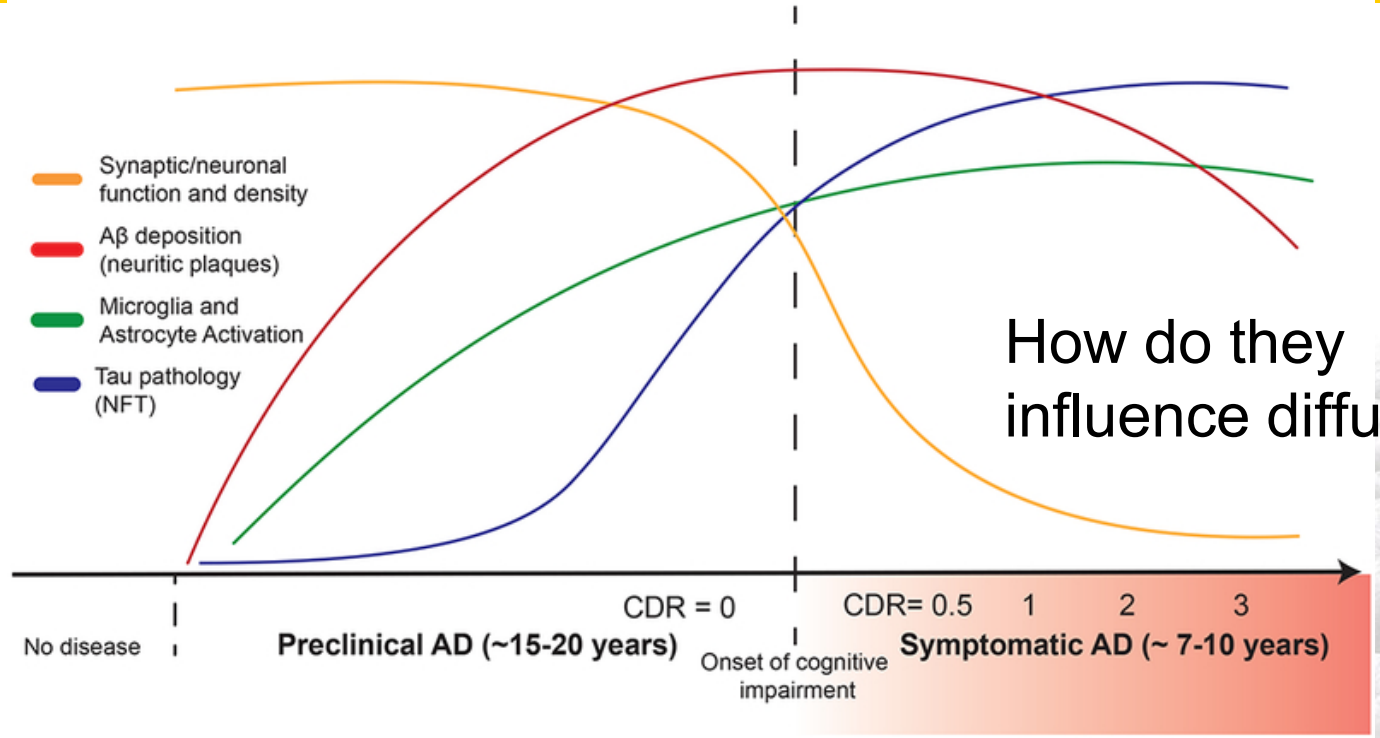
Multiple aspects of microstructure

Mapped across the brain



Nir et al /Jahanshad Mol Psych 2023

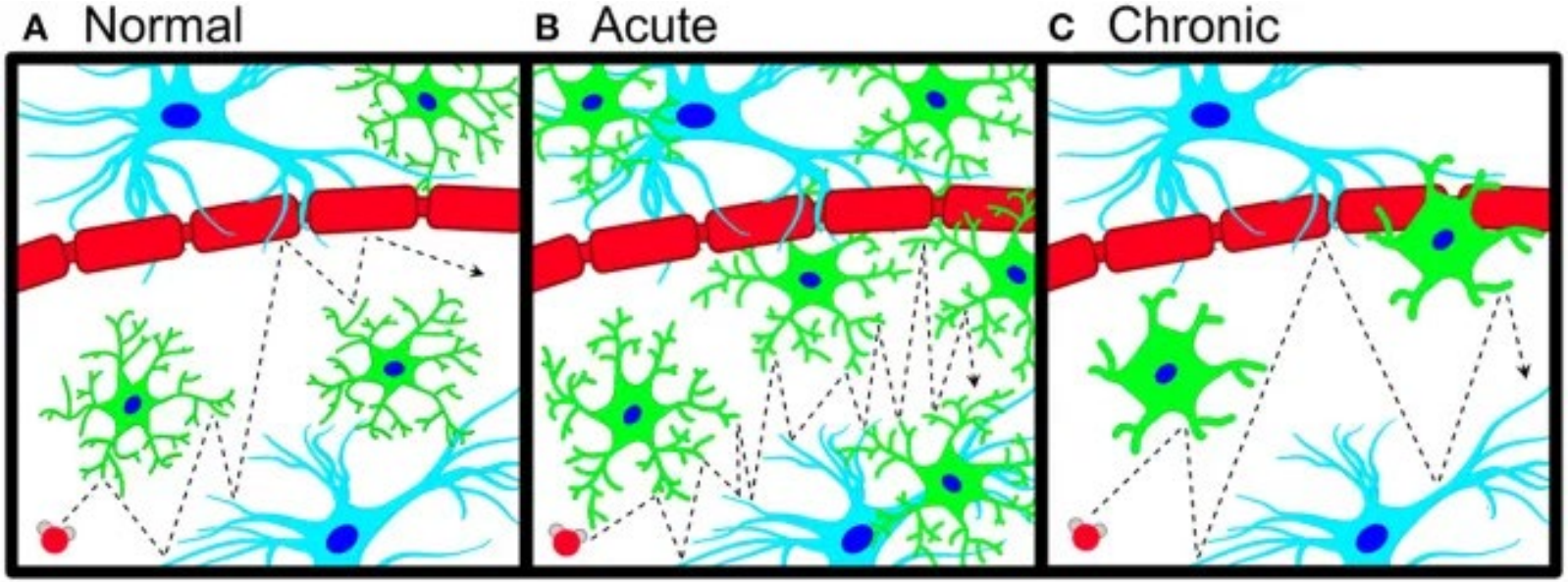
Many pathophysiological processes



How do they influence diffusivity?

Long & Holtzman 2019

Microglial neuroinflammation and MR diffusivity



Yi et al 2019

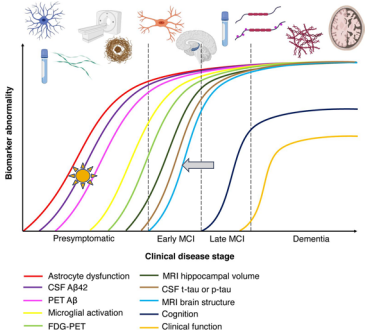
Microglial neuroinflammation and MR diffusivity

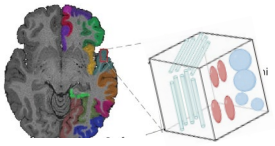
More barriers
Cellular hypertrophy, inflammatory microglia infiltration, hydrophobic plaques

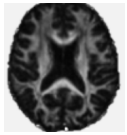
Less barriers
Inflammation edema, less dendritic arborization complexity, degeneration, neurite loss

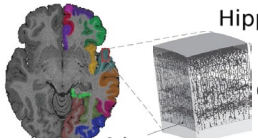
Diffusivity (MD)

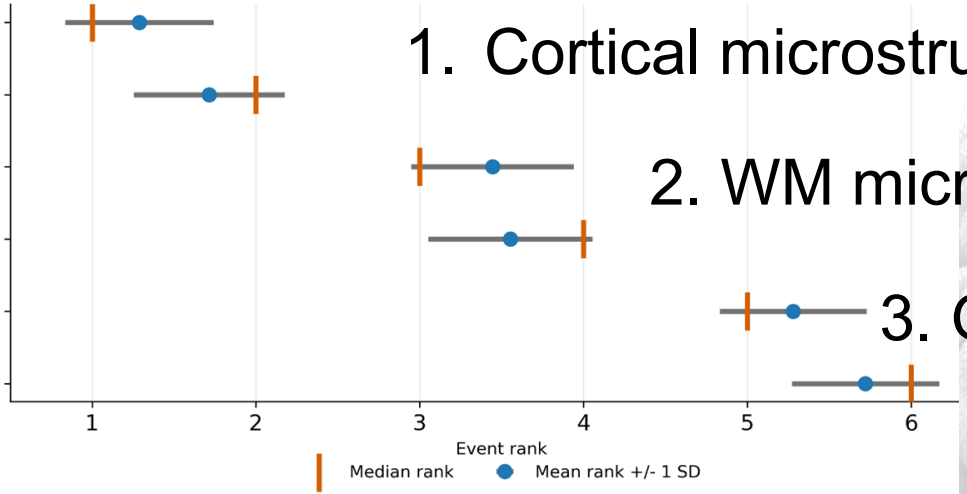
GMm → WMm → Hippocampal vol → GM vol







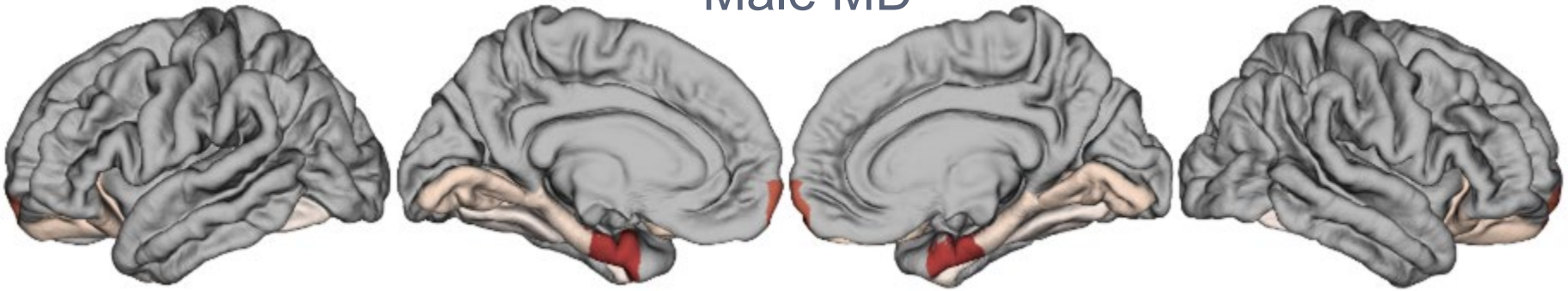




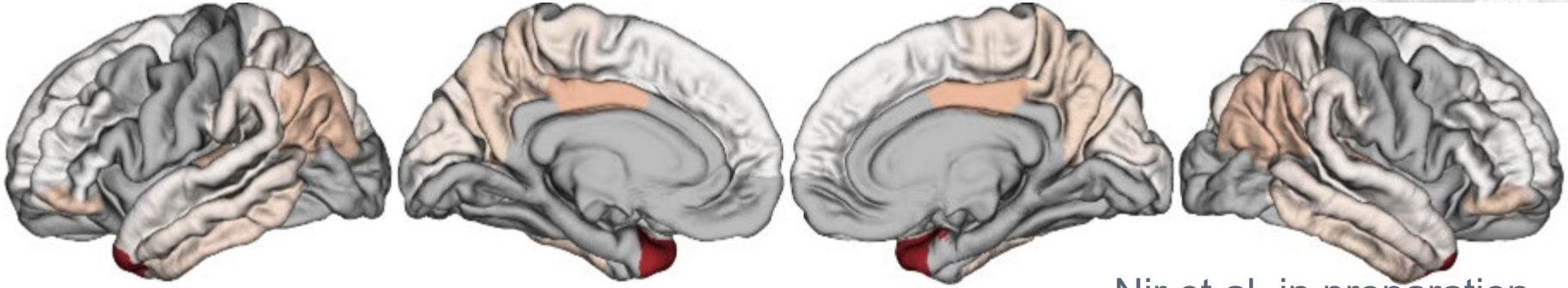
1. Cortical microstructure
2. WM microstructure
3. GM structure

Quadratic beta -amyloid effect in MTL

Male MD

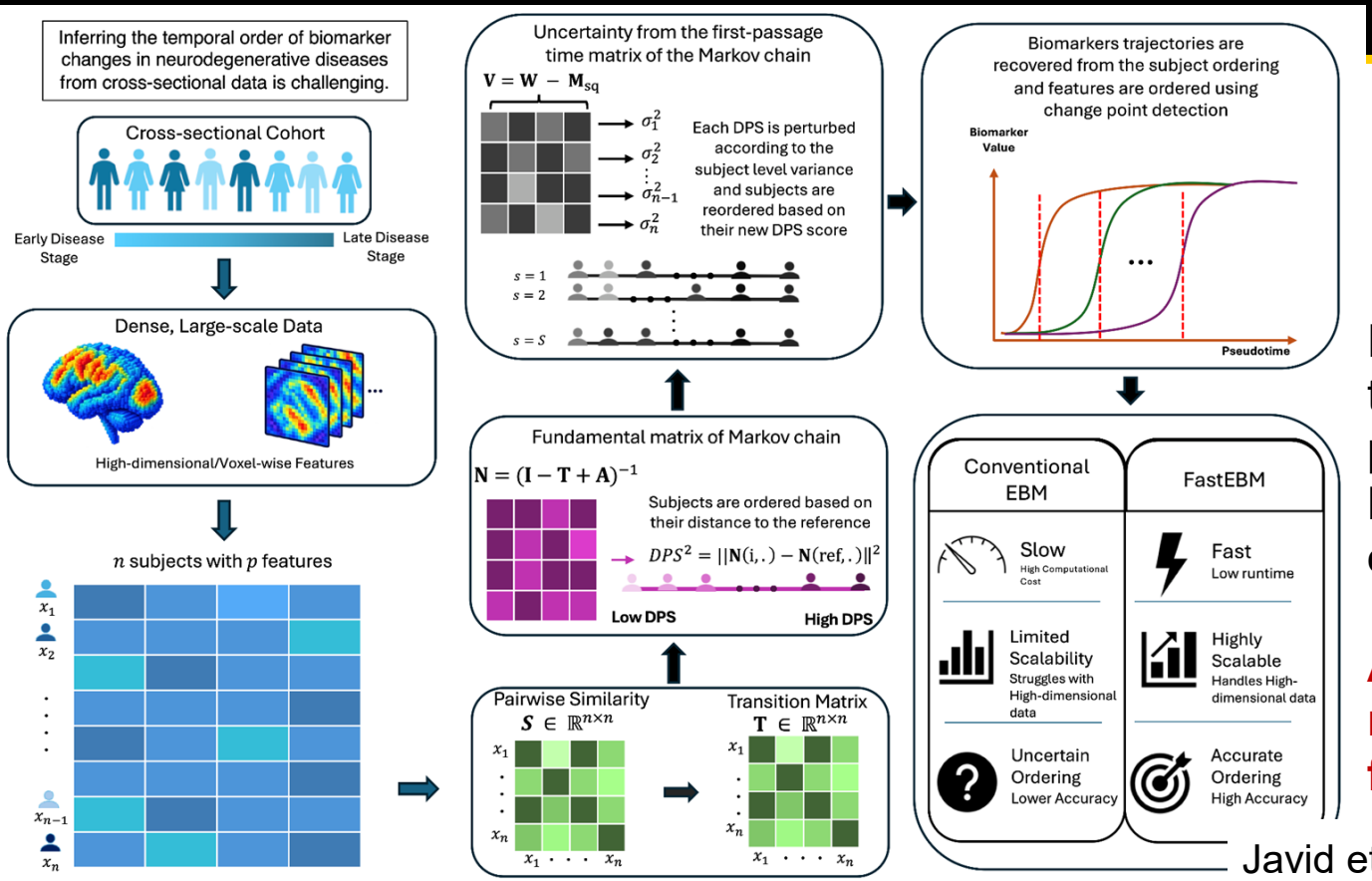


Female MD



Nir et al, in preparation

Fast-EBM: EBM for large feature-sets





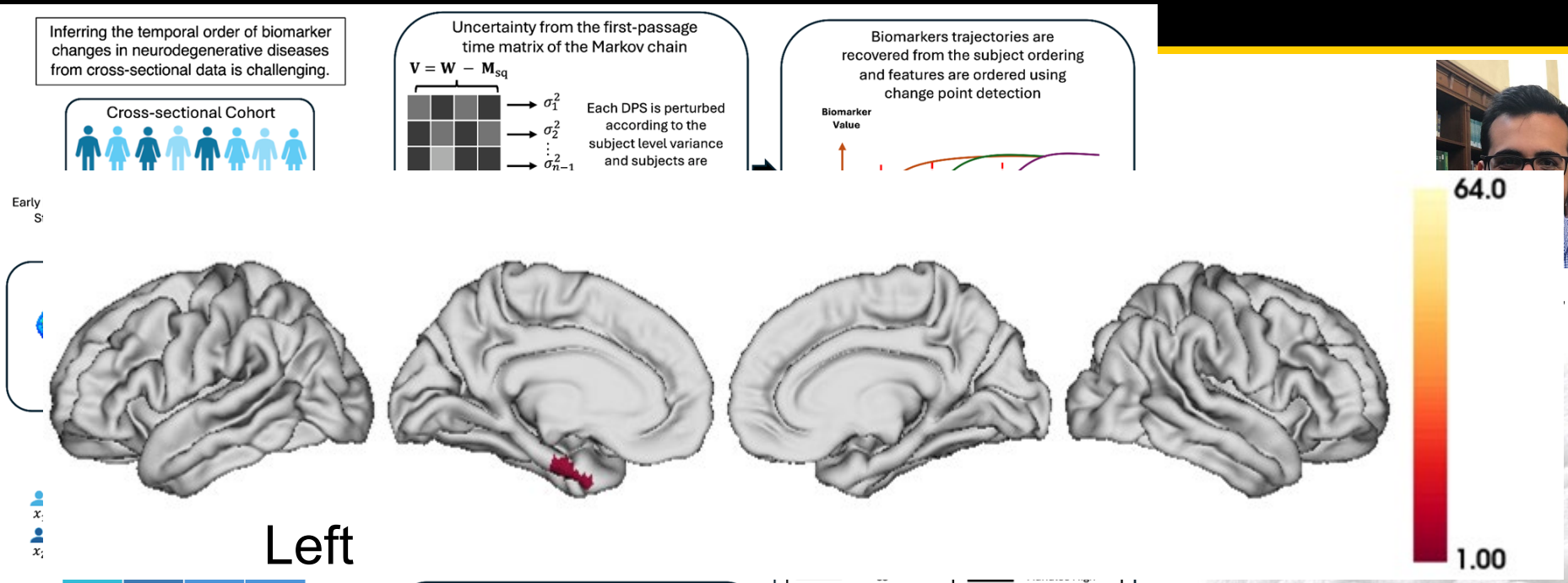
Shayan Javid, MS

FastEBM uses latent-temporal progression-based Bayesian method as opposed to max likelihood

Accurate event ordering no longer limited to ~10 features

Javid et al /Jahanshad biorxiv 2026

Fast-EBM: EBM for large feature-sets

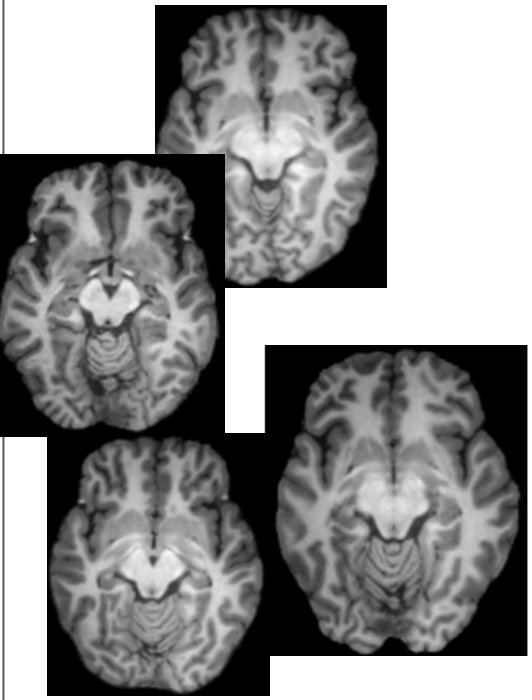


MS

method as opposed to max likelihood

Javid et al /Jahanshad biorxiv 2026





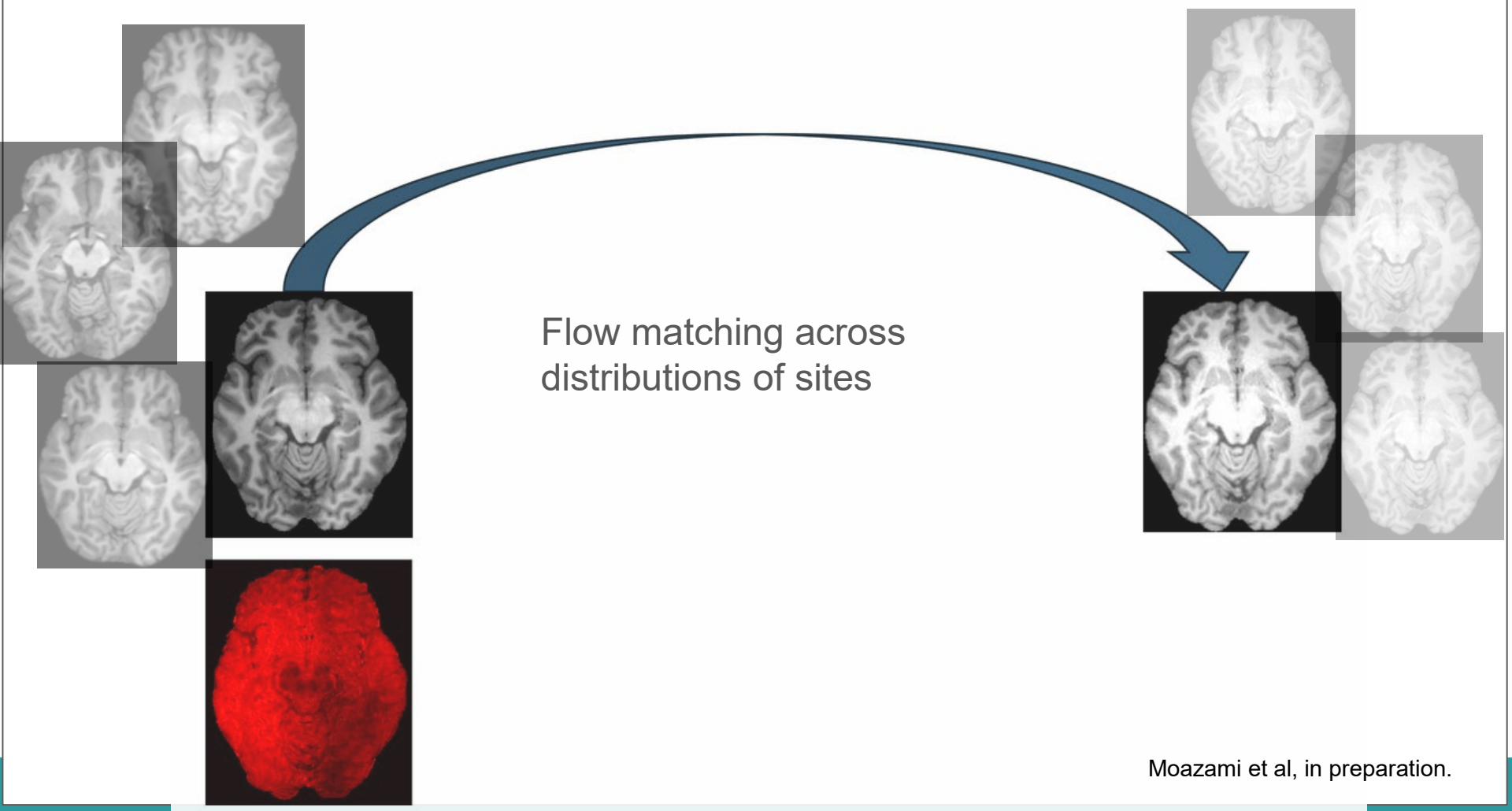
Imaging site 1



Imaging site 2

Saeed Moazami

Moazami et al, in preparation.

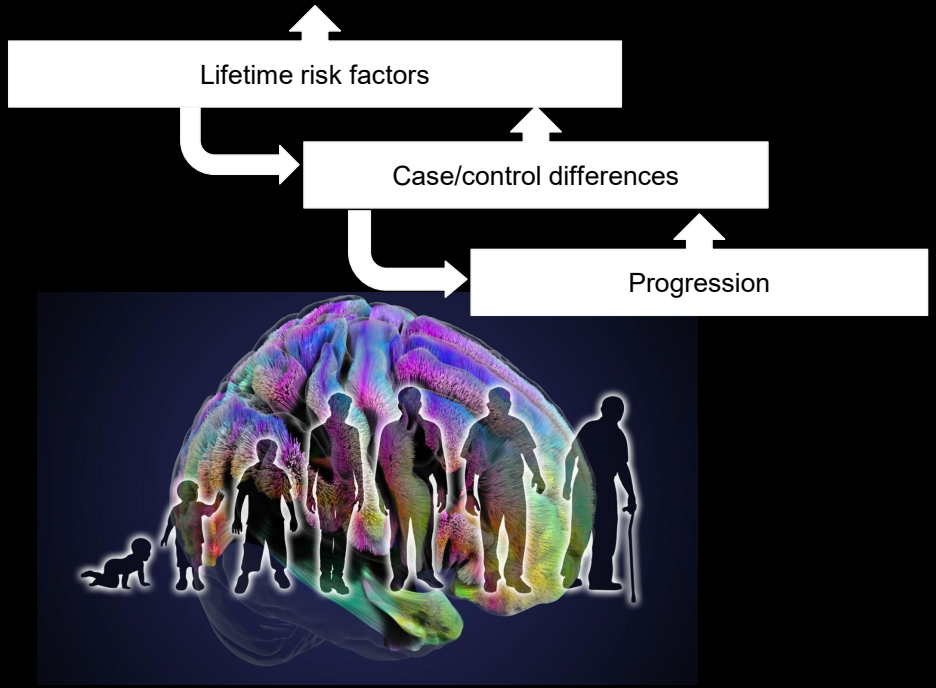


Moazami et al, in preparation.

Ultimate Goal:

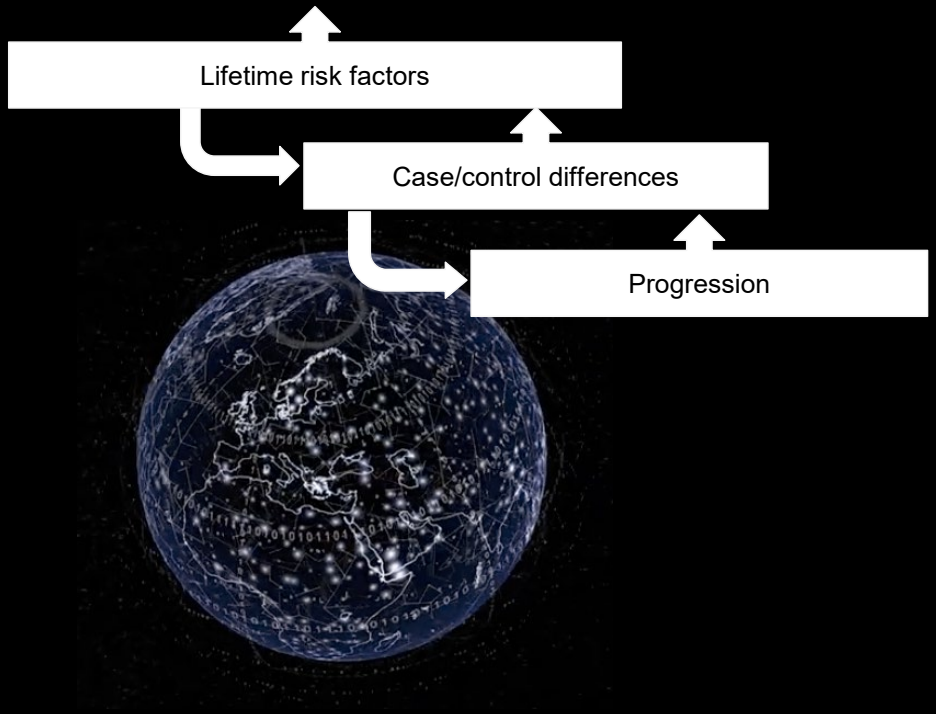
Optimizing individual brain health

Understanding susceptibility to disease and identifying personal preventative strategies and treatments



Neuroimaging + health data + advanced methods
In populations around the world

Understanding susceptibility to disease and identifying personal preventative strategies and treatments



Modeling the individual aging brain



Age: 64



Age: 64

Gadewar et al, / Jahanshad 2023
Moazami et al, in progress



Data used for model training:
200 individuals for each age bin (100M/100F) for training sex-stratified models

Gadewar et al, 2023
Gadewar et al, in progress

1st scan

1 scan

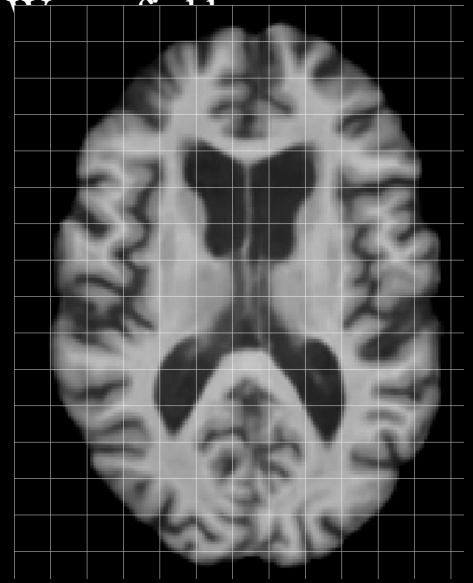


Age: 73



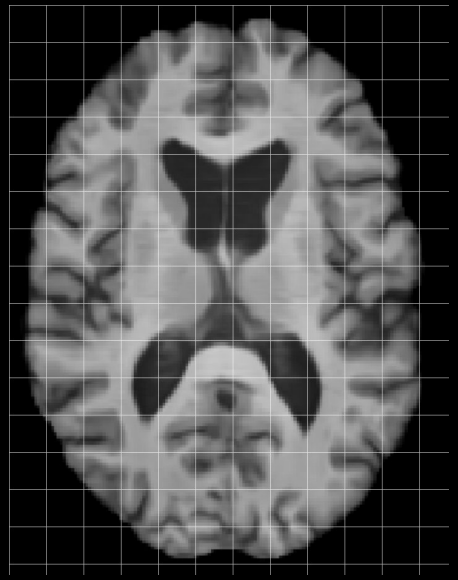
Age: 75

2nd scan

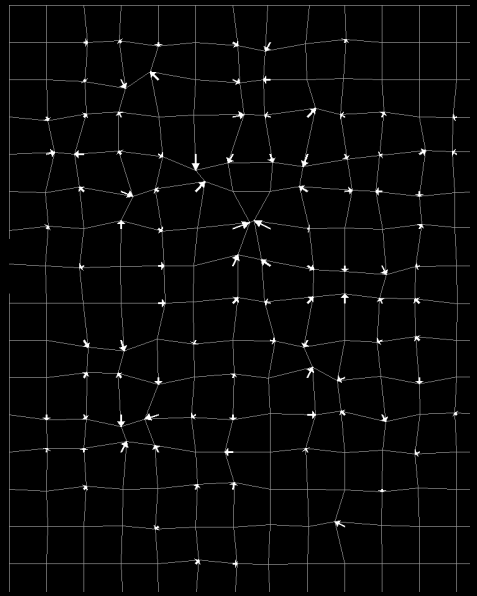


Age: 75

Predicted 2nd scan

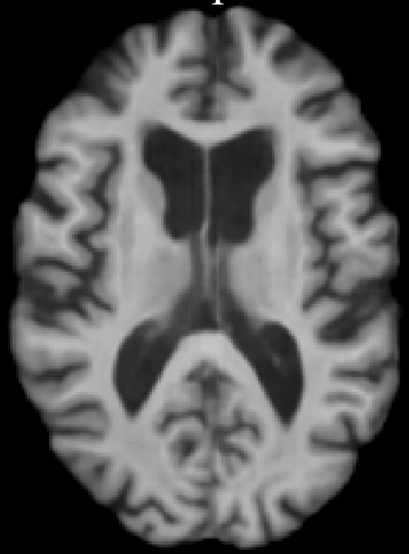


Age: 75



Predicting how your specific brain will age

Individualized warp



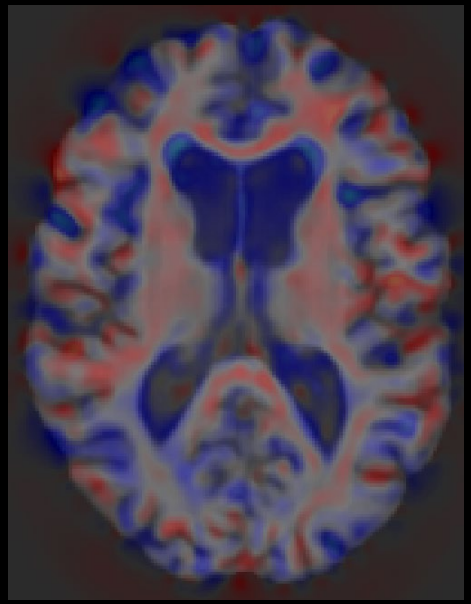
Age: 75

2nd scan



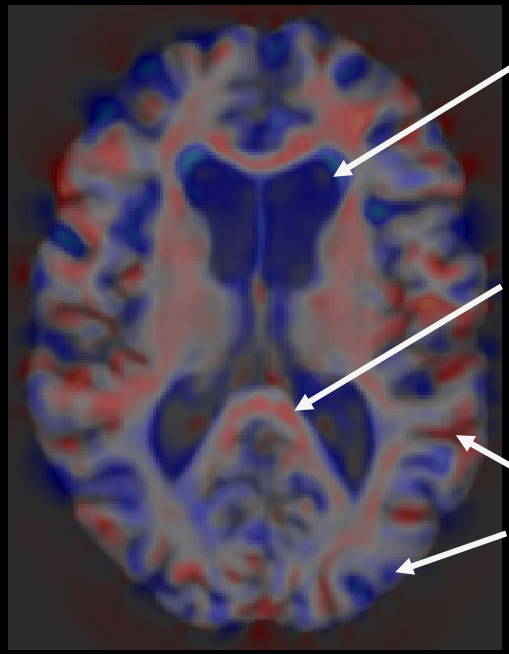
Age: 75

Predicted 2nd scan



Gadewar et al, 2023

What can I do about these deviations?

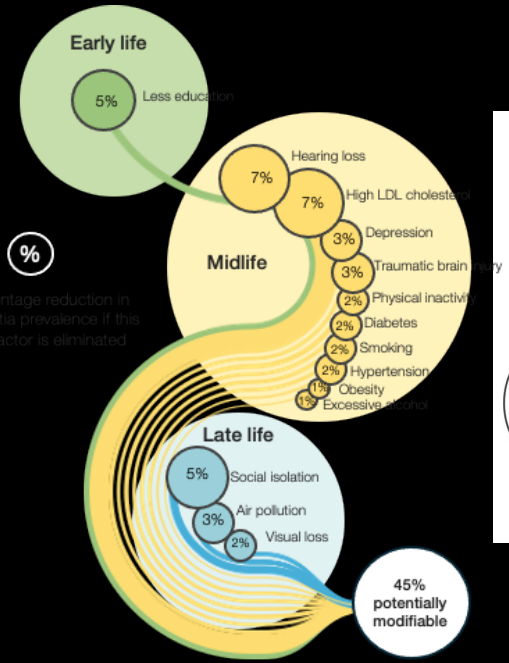


Can I reduce the rate of atrophy here?

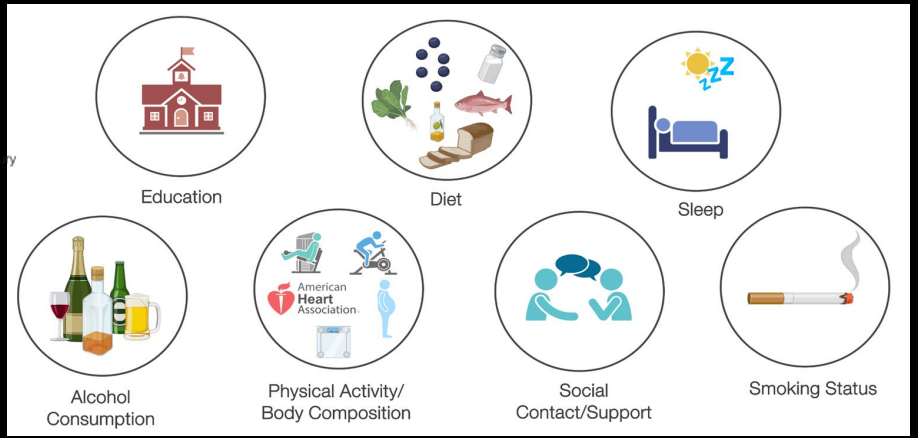
Can I improve white matter integrity here?

Can I improve cortical thickness here, despite my genetics?

What changes can I make in midlife ?
How can they affect my brain?



Livingston et al 2024, Lancet Commission



Age: 45.0



Can we estimate *causal* changes to regional brain metrics?

In the absence of clinical trials, can we model causal effects using large epidemiological studies? Causal inference models can be used to predict counterfactuals (alternate outcomes) similar to properly randomized experiments.



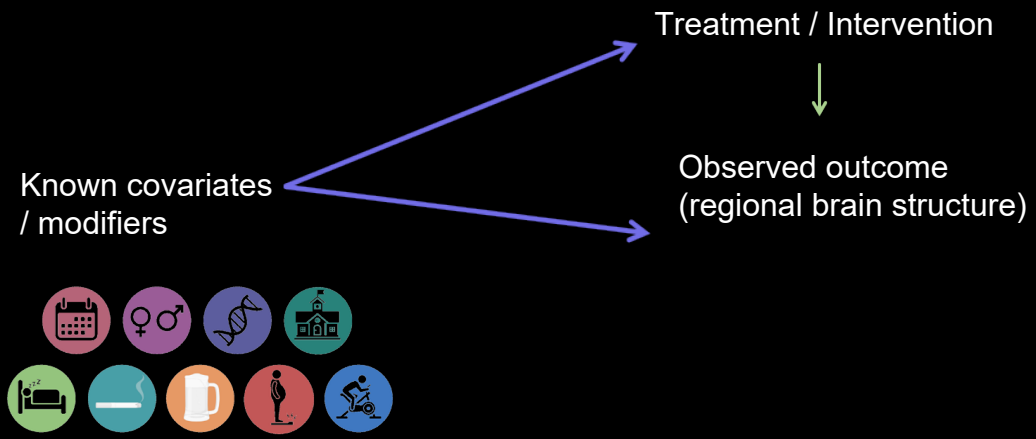
Myrl Marmarelis, PhD

Treatment / Intervention

↓

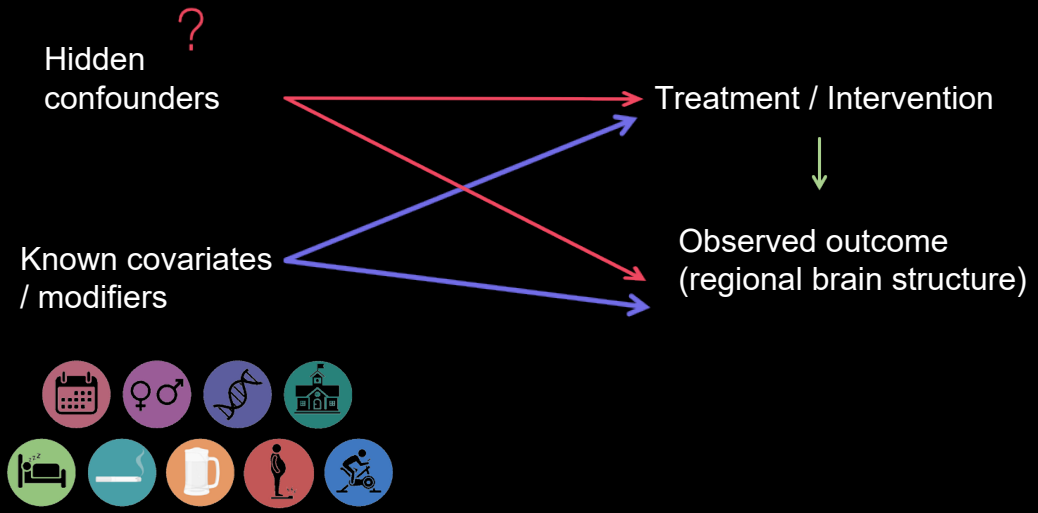
Observed outcome
(regional brain structure)

Can we estimate *causal* changes to regional brain metrics?



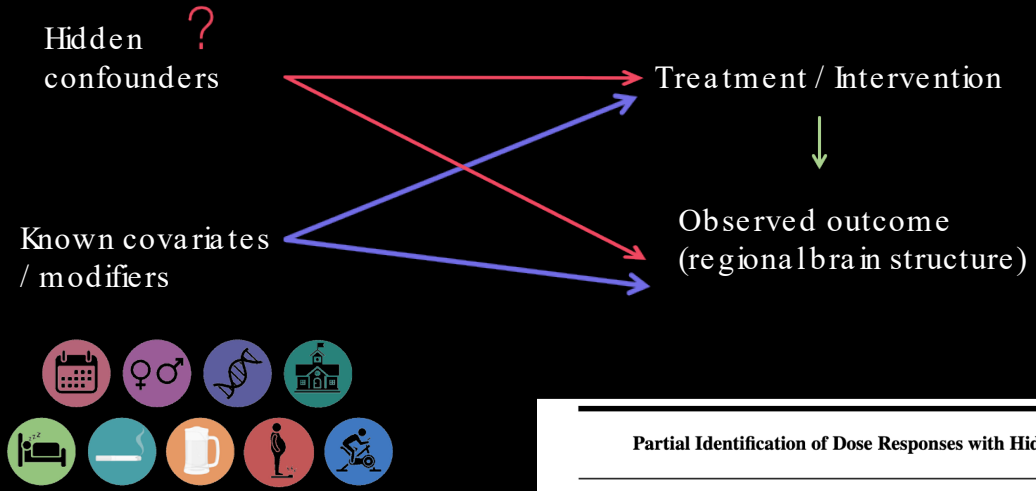
Can we estimate *causal* changes to regional brain metrics?

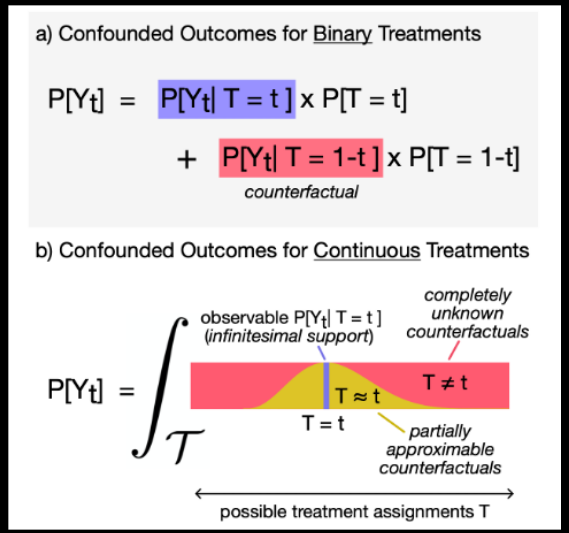
But, there are many unknowns!



Can we estimate *causal* changes to regional brain metrics?

- Causal sensitivity analysis can make reasonable structural assumptions about different levels of hidden confounding
  - Does the relationship still hold at increasing levels of confounding?
- Now developed for continuous treatments, allowing for the quantification of dose responses (Marmarelis et al., 2023)





Partial Identification of Dose Responses with Hidden Confounders

Myrl G. Marmarelis[1] Elizabeth Haddad[2] Andrew Jesson[3]
Neda Jahanshad[2] Aram Galstyan[1] Greg Ver Steeg[1]

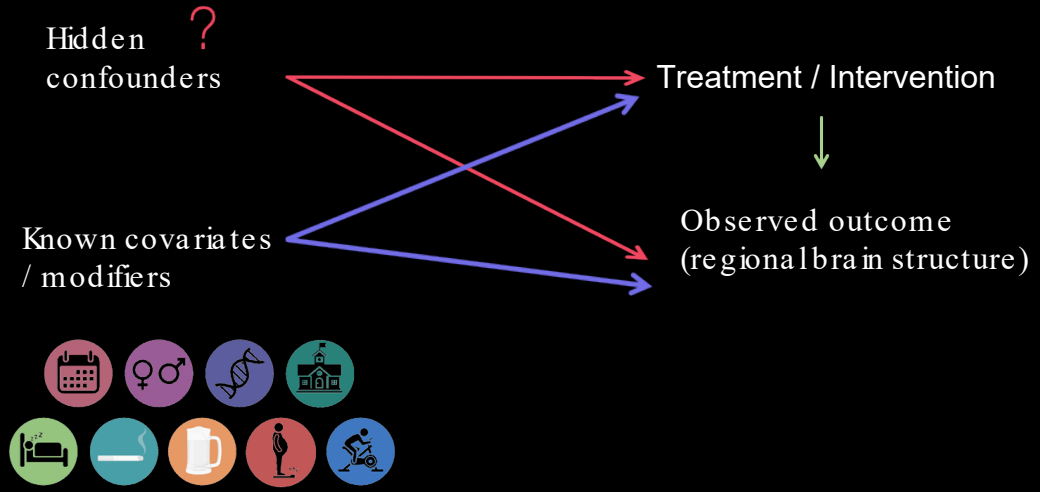
[1]USC Information Sciences Institute, 4676 Admiralty Way, Marina del Rey, CA 90292
[2]USC Stevens Neuroimaging and Informatics Institute, 4676 Admiralty Way, Marina del Rey, CA 90292
[3]University of Oxford, OATML, 14 Parks Road, Oxford, UK OX1 3AQ

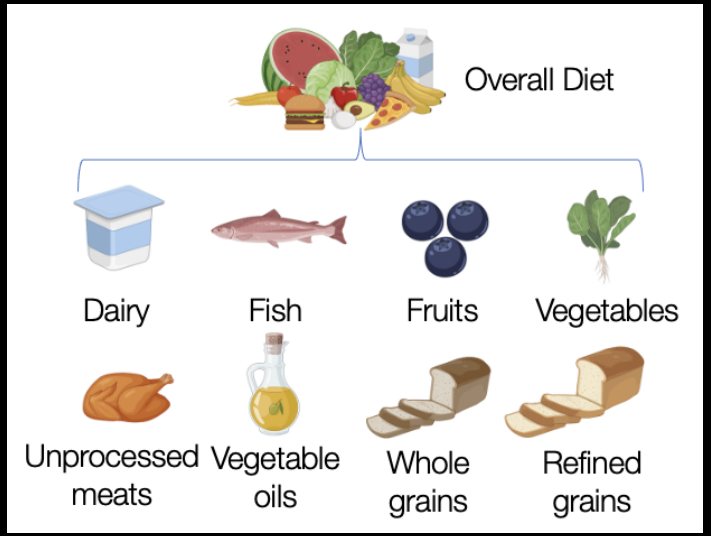
Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023), PMLR 216:1368–1379.

Marmarelis et al2023

What specific improvements can I make?

- Can an incremental improvement in one dietary aspect improve my brain health?



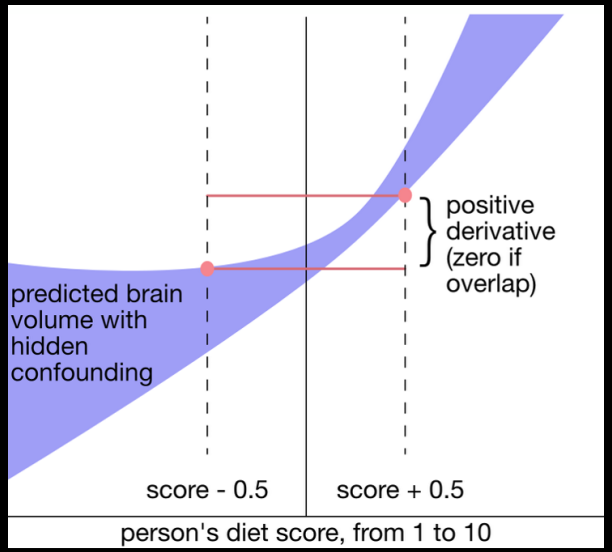


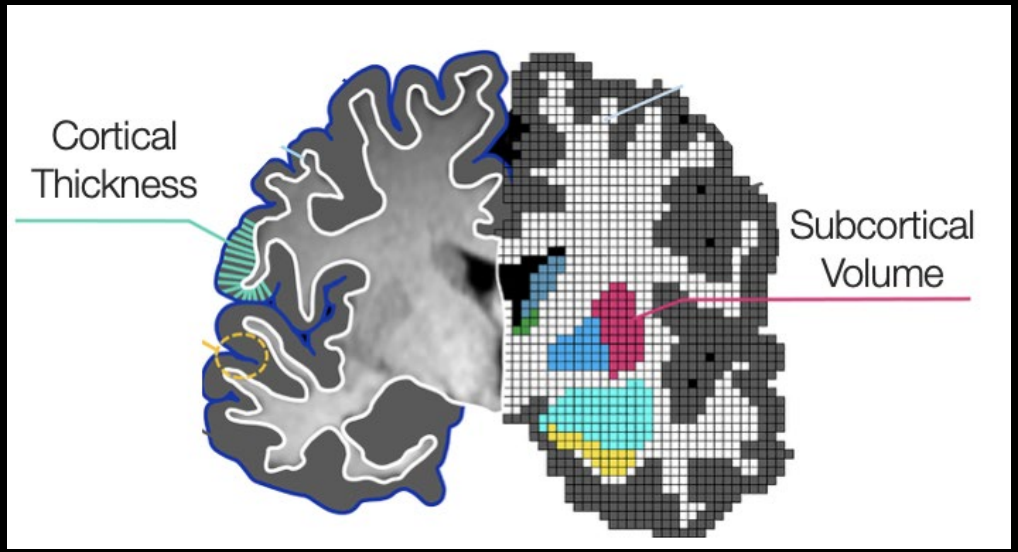
Haddad, Marmarelis et al 2023

Causal sensitivity analysis - causal dietary effects on brain structure

Framed as: “a one unit increase in a specific diet score can cause an X% increase in cortical thickness”

Different confounding levels will have different prediction variations to assess effect robustness

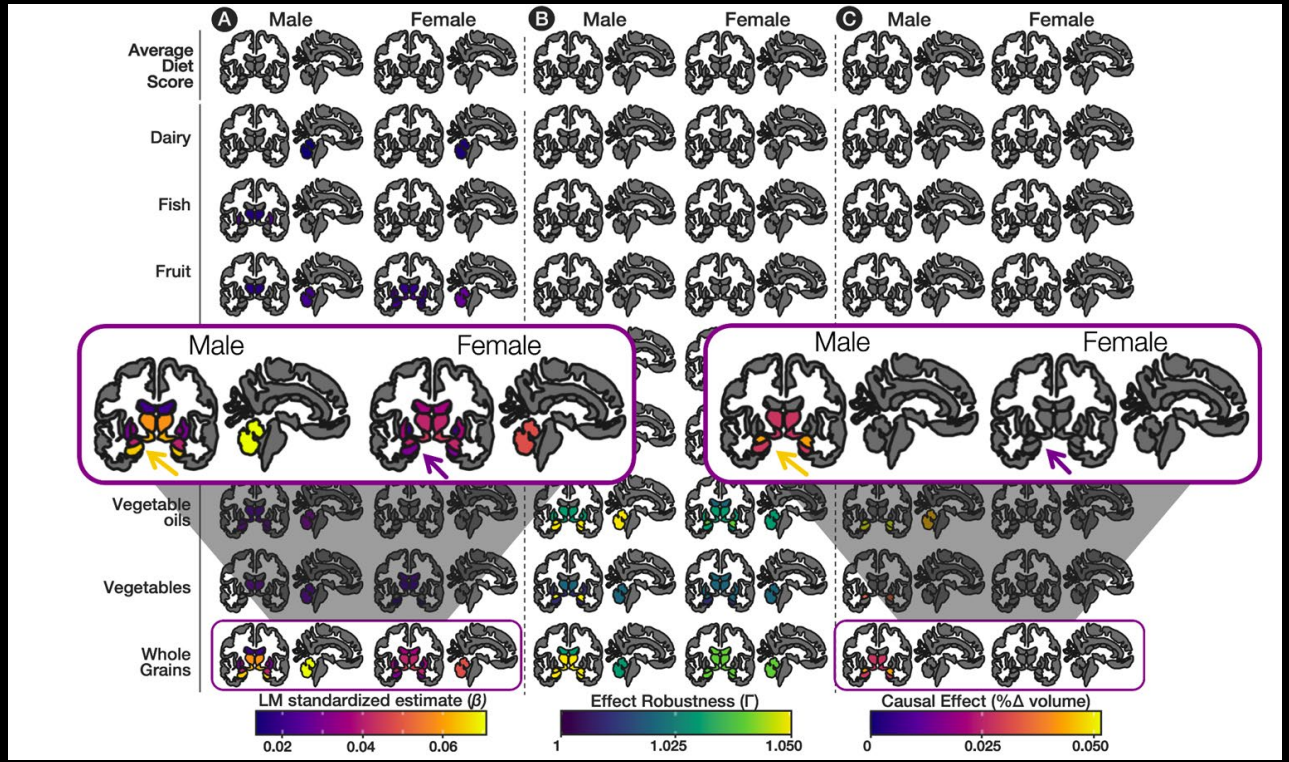


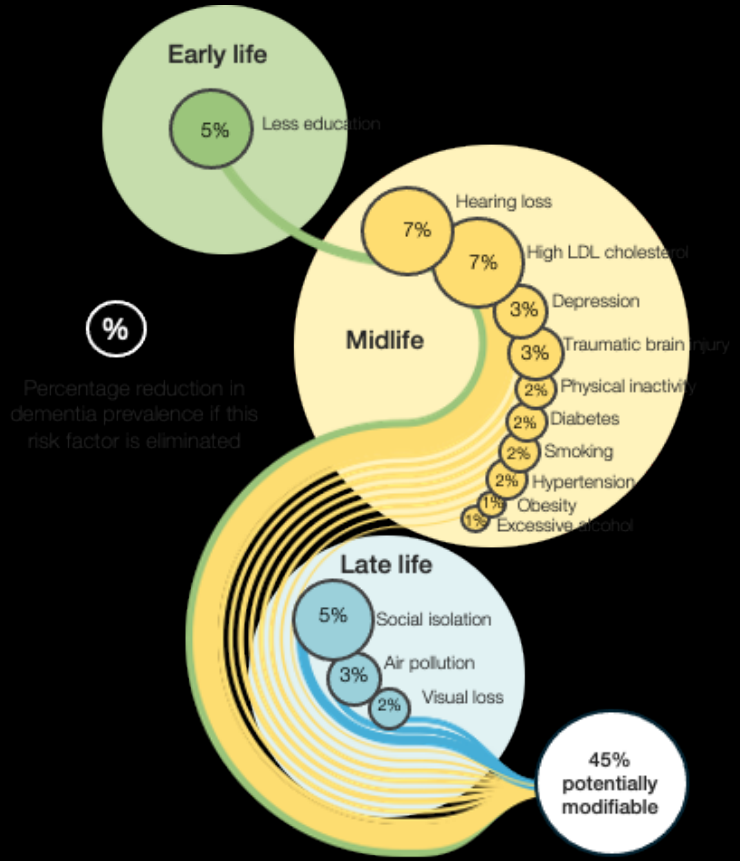


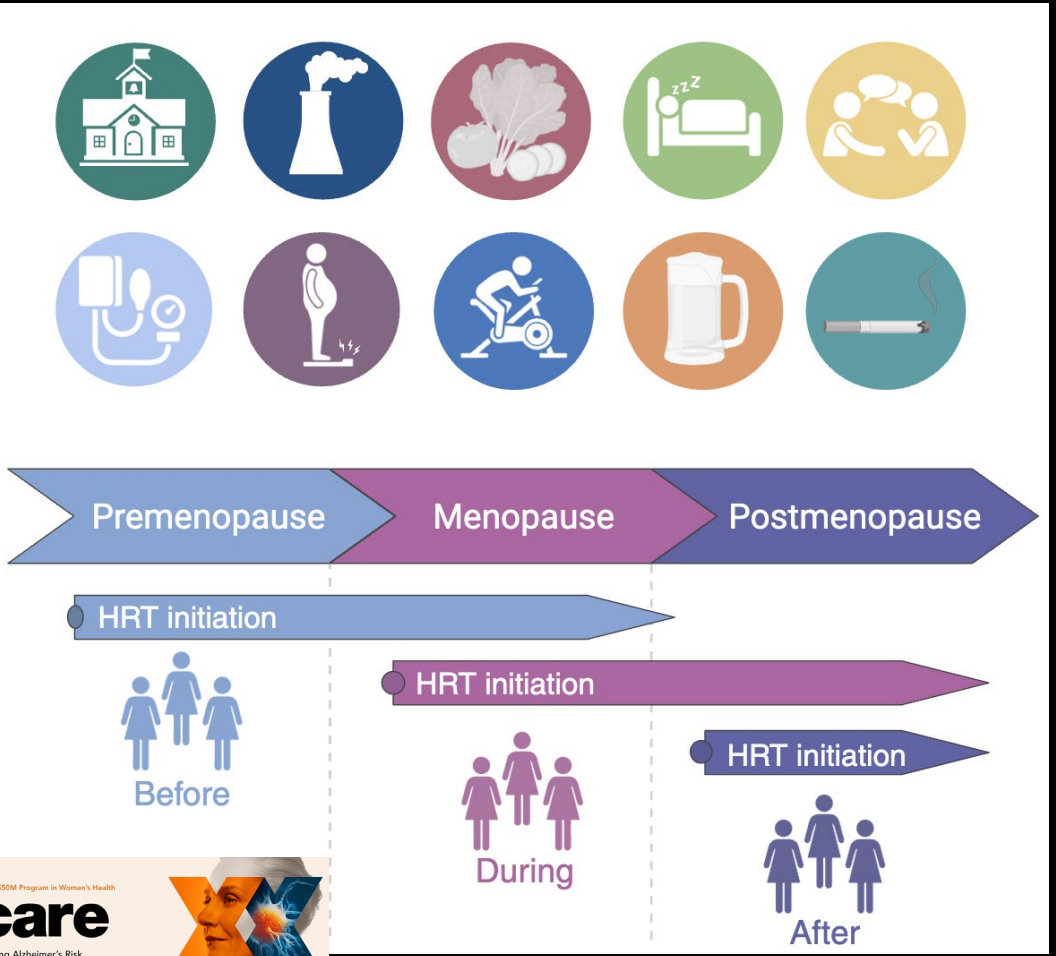
Haddad, Marmarelis et al 2023

Causal sensitivity analysis reveals more than associations alone

Averaged dietary effects across males and females separately, no causal relationship for females




Early life
5%
Less education
Midlife
7%
Hearing loss
7%
High LDL cholesterol
3%
Depression
3%
Traumatic brain injury
2%
Physical inactivity
2%
Diabetes
2%
Smoking
2%
Hypertension
1%
Obesity
1%
Excessive alcohol
%
Percentage reduction in dementia prevalence if this risk factor is eliminated
Late life
5%
Social isolation
3%
Air pollution
2%
Visual loss
45% potentially modifiable


Premenopause
Menopause
Postmenopause
HRT initiation
Before
HRT initiation
During
HRT initiation
After

NEW $50M Program in Women's Health
care
Cutting Alzheimer's Risk through Endocrinology