Supplementary Information for: **Grey-matter structure markers of Alzheimer's disease, Alzheimer's conversion, functioning and cognition: a meta-analysis across 11 cohorts**

Baptiste Couvy-Duchesne[1,2], Vincent Frouin[3], Vincent Bouteloup[4], Nikitas Koussis[5,6], Julia Sidorenko[1], Jiyang Jiang[7], Alle Meije Wink[8,9], Luigi Lorenzini[8,9], Frederik Barkhof[8,10], Julian N. Trollor[11], Jean-François Mangin[3], Perminder S. Sachdev[7,12], Henry Brodaty[7], Michelle K. Lupton[13,14,15], Michael Breakspear[5,6], Olivier Colliot[2,*], Peter M. Visscher[1,16,*], Naomi R. Wray[1,17,*], for the Alzheimer's Disease Neuroimaging Initiative‡, the Australian Imaging Biomarkers and Lifestyle flagship study of ageing†, the Alzheimer's Disease Repository Without Borders Investigators¥, and the MEMENTO cohort Study Group§.

* these authors contributed equally.

Correspondence: BCD (b.couvyduchesne@uq.edu.au)

This file includes:

Supplementary Tables 3
STable 1: Description of the samples included in the analyses. 4
Supplementary Figures 6
SFigure 1: Concordance of means (top panels) and variances (bottom panels) of the ROI based measurements 8
SFigure 2: Concordance of means (top panels) and variances (bottom panels) of the grey-matter vertices of cortical thickness 9
SFigure 3: Concordance of means (top panels) and variances (bottom panels) of the grey-matter vertices of cortical surface area 10
SFigure 4: Concordance of means (top panels) and variances (bottom panels) of the grey-matter vertices of subcortical thickness. 11

SFigure 5: Concordance of means (top panels) and variances (bottom panels) of the grey-matter vertices of subcortical surface area 12
SFigure 6: Concordance of means (top panels) and variances (bottom panels) of the grey-matter vertices of cortical surface area – after removing "noisy vertices". 13
SFigure 7: Position of the noisy vertices for cortical thickness (L, R), cortical surface area (L, R), subcortical thickness (L, R) and subcortical surface area (L, R). 14
SFigure 8: Vertex-wise morphometricity estimated using two different mixed models. .. 15
SFigure 9: Forest plots for the meta-analysis of vertex-wise morphometricity – disease status, conversion and family history 16
SFigure 10: Forest plots for the meta-analysis of vertex-wise morphometricity – neuropsychological scales 18
SFigure 11: Forest plots of vertex-wise morphometricity using a mixed models with 4 variance components 20
SFigure 12: Morphometricity estimates of neuropsychological scales without controlling for disease status 22
SFigure 13: Forest plots of the marginal (multivariate) and univariate association between left hippocampal volume and RAVLT delayed recall score 23
SFigure 14: Univariate ROI associations with Alzheimer's conversion within 3 years of imaging 24
SFigure 15: Univariate ROI associations with memory scores 25
SFigure 16: Univariate ROI associations with neurological scales 26
SFigure 17: Number of significant clusters (after Bonferroni correction) of the different models of vertex-wise associations 27
SFigure 18: Unthresholded map of association (betas) for AD vs. HC, using the GLM approach. 28
SFigure 19: Concordance of brain regions associated with Alzheimer's disease (AD vs. HC) across the different analyses (ROI and vertex-wise), by type of measurement. 29
SFigure 20: Concordance of brain regions associated with Alzheimer's conversion (at 3 years) across the different analyses. 30
SFigure 21: Concordance of brain regions associated with Functional Activities Questionnaire (FAQ) score across the different analyses. 31
SFigure 22: Concordance of brain regions associated with Logical Memory (delayed recall) score across the different analyses. 32
SFigure 23: Concordance of brain regions associated with MMSE score across the different analyses. 33
SFigure 24: Concordance of brain regions associated with RAVLT immediate memory score across the different analyses. 33
SFigure 25: Concordance of brain regions associated with RAVLT delayed memory score across the different analyses. 34
SFigure 26: All associated vertex-wise measurements in any of the 24 traits considered. 35
SFigure 28: Comparison of ROI-based morphometricity estimates between the discovery and replication samples. 37
Supplementary Tables (Captions) 38
Appendices 41
Appendix A. Sample description, including acquisition protocols of MRI images 42
Appendix B. Conversion of effect sizes – from mean differences to cohen's d. 52

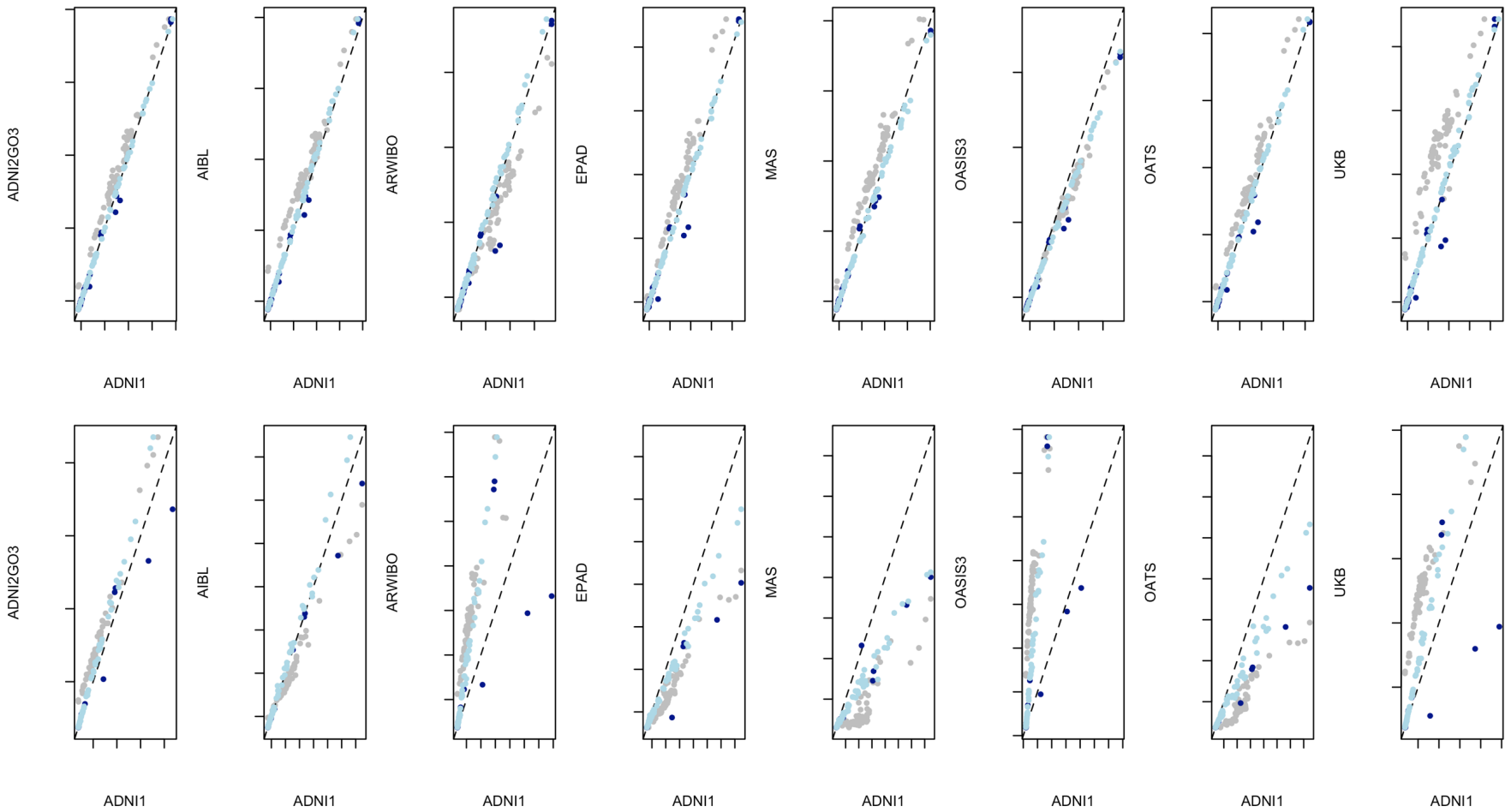
Supplementary Tables

STable 1: Description of the samples included in the analyses.

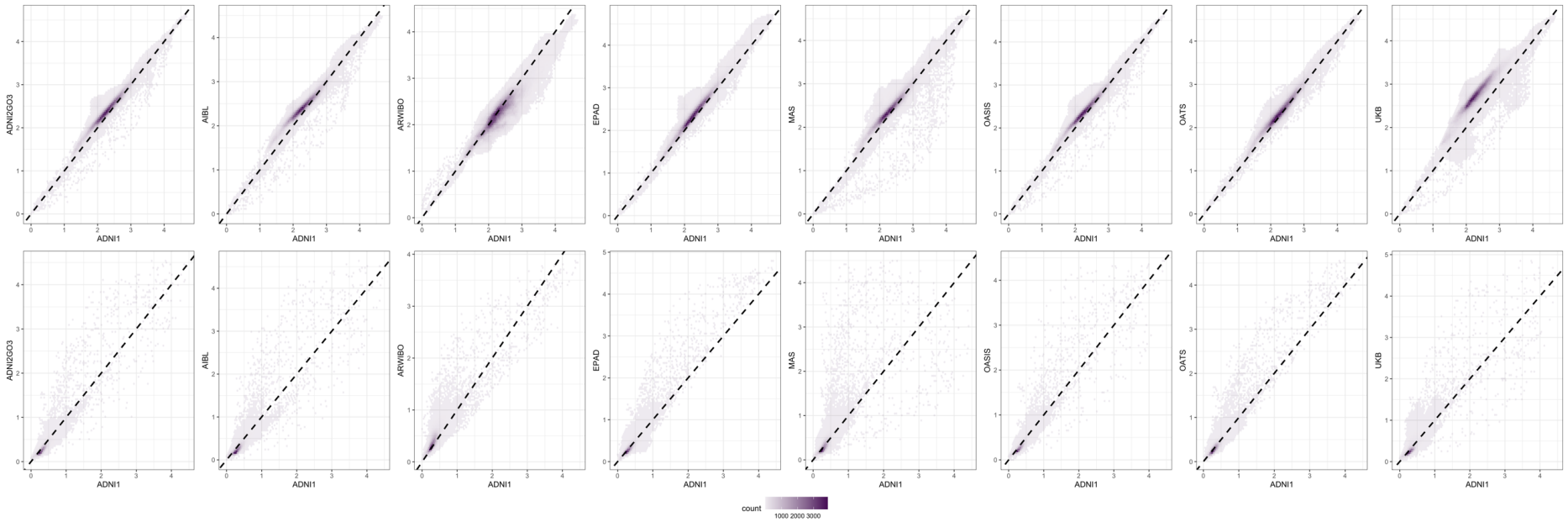
		ADNI1	ADNI2GO3	AIBL	ARWIBO	EPAD	MAS	OASIS3	OATS	UKB	MEMENTO	PISA
N		**805**	**1,410**	**606**	**934**	**1,315**	**527**	**1,019**	**365**	**37,644**	**1,880**	**279**
Age	Mean (SD)	75.2 (6.8)	72.1 (7.19)	72.8 (6.7)	56.7 (16.0)	64.4 (7.1)	78.4 (4.7)	70.7 (9.3)	70.3 (5.1)	63.6 (7.5)	70.0 (8.7)	60.6 (6.9)
Sex	N female (%)	465 (57.8%)	733 (50.3%)	335 (55.3%)	565 (60.5%)	748 (56.8%)	290 (55.0%)	565 (55.4%)	240 (65.8%)	20,056 (53.3%)	1,199 (63.8%)	193 (72.3%)
AD	N (%)	186 (23.0%)	195 (13.8%)	72 (11.9%)	130 (13.9%)	0 (0%)	0 (0%)	207 (20.3%)	6 (1.4%)	NA	0 (0%)	26 (9.7%)
MCI	N (%)	393 (49.0%)	619 (43.9%)	89 (14.7%)	0 (0%)	0 (0%)	175 (33.2%)	21 (0.02%)	46 (12.6%)	NA	1,575 (83.8%)	21 (7.9%)
HC	N (%)	226 (28.0%)	577 (41.0%)	443 (73.1%)	715 (76.6%)	1,315 (100%)	288 (54.6%)	791 (77.6%)	308 (84.4%)	NA	305 (16.2%)	220 (82.4%)
conversion 1 year	N conv. (%); N non-conv.	48 (17.4%); 227	34 (6.9%); 457	2 (3.8%); 51	NA	7 (0.8%); 787	NA	20 (4.1%); 471	NA	NA	22 (1.4%); 1,555 (98.6%)	NA
2 years		122 (36.7%); 210	64 (14.5%); 376	2 (4.9%); 39	NA	9 (3.6%); 238	4 (1.1%); 369	39 (9.1%); 388	1 (0.03%); 263	2 (<0.01%); 37,642	72 (4.7%); 1,467 (95.3%)	NA
3 years		153 (46.7%); 174	98 (29.3%); 236	3 (9.4%); 29	NA	9 (12.3%); 74	6 (1.6%); 358	51 (13.8%); 319	2 (0.7%); 256	2 (<0.01%); 37,642	122 (8.2%); 1,373 (91.8%)	NA
4 years		177 (58.6%); 125	114 (34.7%); 215	5 (20.8%); 19	NA	NA	20 (0.058%); 325	63 (20.4%); 246	3 (1.5%); 197	5 (<0.01%); 37,639	164 (11.3%); 1,282 (88.6%)	NA
5 years		195 (63.9%); 110	126 (42.0%); 174	5 (38.5%); 8	NA	NA	25 (7.5%); 309	73 (28.3%); 185	5 (5.1%); 94	5 (<0.01%); 37,639	201 (18.6%); 881 (81.4%)	NA
10 years		232 (84.0%); 44	NA	NA	NA	NA	NA	NA	NA	6 (<0.01%); 37,638	NA	NA
MMSE	mean (SD); N	26.7 (2.7); 805	27.7 (2.6); 1410	27.5 (3.6); 606	27.0 (4.1); 933	28.6 (1.7); 1310	28.1 (1.4); 527	27.9 (2.9); 1,017	28.4 (1.6); 361	NA	28.0 (1.85); 1,875	NA
CDR	mean (SD); N	0.42 (0.30); 805	0.33 (0.31); 1410	0.21 (0.39); 606	0.18 (0.42; 817)	0.094 (0.19); 1307	0.074 (0.18); 527	0.19 (0.34); 1,019	NA	NA	0.29 (0.25); 1,869	NA
FAQ	mean (SD); N	4.99 (6.60); 802	3.20 (5.70); 1386	NA	NA	NA	NA	2.23 (4.8); 922	NA	NA	NA	NA
GDS	mean (SD); N	1.39 (1.35); 805	1.34 (1.44); 1409	NA	NA	1.63 (2.01); 1308	NA	1.68 (2.08); 921	1.59 (1.89); 354	NA	NA	NA
NPI-Q	mean (SD); N	1.83 (2.77); 805	1.23 (2.19); 1391	NA	NA	NA	NA	1.23 (1.95); 921	0.54 (1.47); 283	NA	NA	NA
RAVLT - learning	mean (SD); N	3.65 (2.67); 801	4.85 (2.75); 1404	NA	NA	NA	5.73 (2.34); 523	NA	5.71 (2.22); 338	NA	NA	5.59 (2.35); 279
RAVLT – immediate	mean (SD); N	32.4 (11.4); 801	38.8 (12.9); 1404	NA	NA	NA	40.7 (9.77); 523	NA	44.1 (9.30); 338	NA	NA	46.2 (12.4); 279
RAVLT – delayed	mean (SD); N	3.61 (3.98); 804	5.67 (4.67); 1404	NA	NA	NA	7.54 (3.54); 522	NA	8.46 (3.25); 337	NA	NA	8.8 (4.1); 279
RAVLT - forgetting	mean (SD); N	4.32 (2.37); 801	4.24 (2.68); 1400	NA	NA	NA	2.95 (2.21); 522	NA	2.66 (2.10); 337	NA	NA	2.6 (2.3); 279

RAVLT - % forgetting	mean (SD); N	63.3 (34.7); 797	51.7 (34.2); 1398	NA	NA	NA	30.8 (24.6); 521	NA	25.21 (20.90); 337	NA	NA	27.4 (29.3); 279
Logical memory - delayed	mean (SD); N	5.78 (5.41); 805	8.88 (5.17); 1407	8.97 (5.49); 606	NA	NA	9.33 (4.22); 526	9.73 (5.8); 736	10.3 (4.00); 361	NA	NA	NA
Logical memory - immediate	mean (SD); N	8.27 (4.86); 805	10.73 (4.78); 1409	10.8 (5.0); 606	NA	NA	11.04 (4.08); 526	11.2 (5.2); 736	11.77 (3.82); 360	NA	NA	NA
Maternal AD	N (%)	159 (24.9%)	474 (38.7%)	NA	103 (12.2%)	630 (48.2%)	NA	362 (40.5%)	NA	6,712 (19.5%)	NA	NA
Paternal AD	N (%)	53 (8.9%)	201 (16.1%)	NA	54 (6.4%)	302 (23.1%)	NA	183 (20.9%)	NA	3,590 (11,2%)	NA	NA

Supplementary Figures

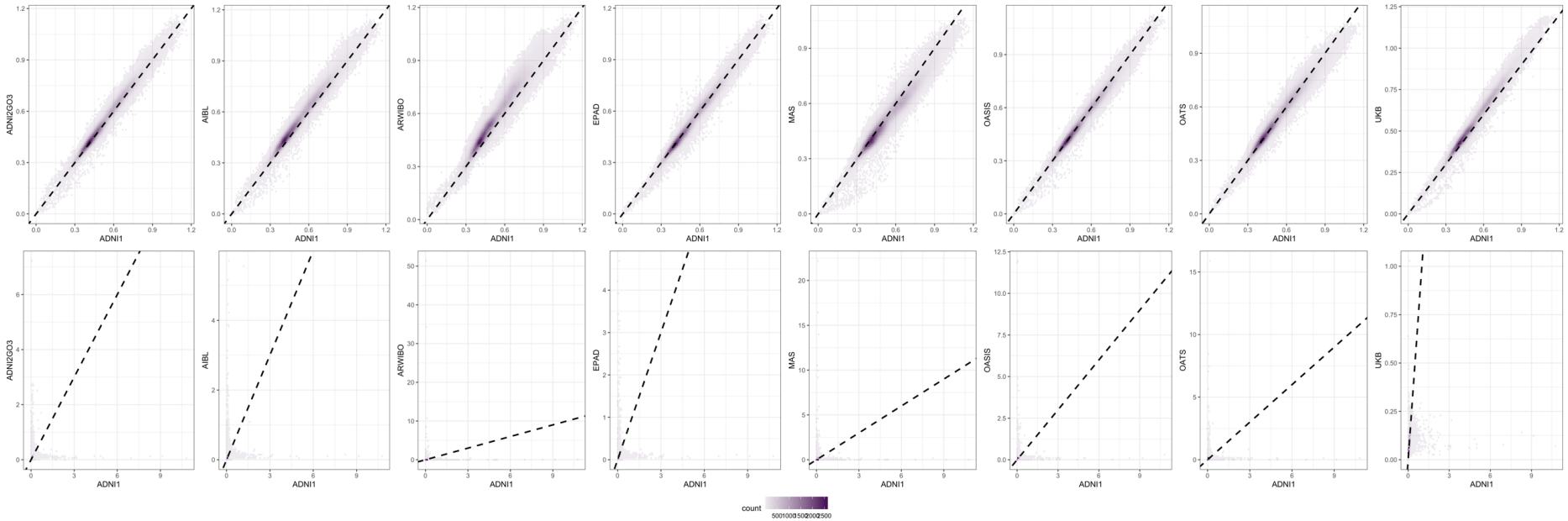


SFigure 1: Concordance of means (top panels) and variances (bottom panels) of the ROI based measurements.
Cortical thickness is shown in grey, surface area in light blue and volumes in dark blue. The variance of ROI based measurement varied in some samples but remained highly correlated.



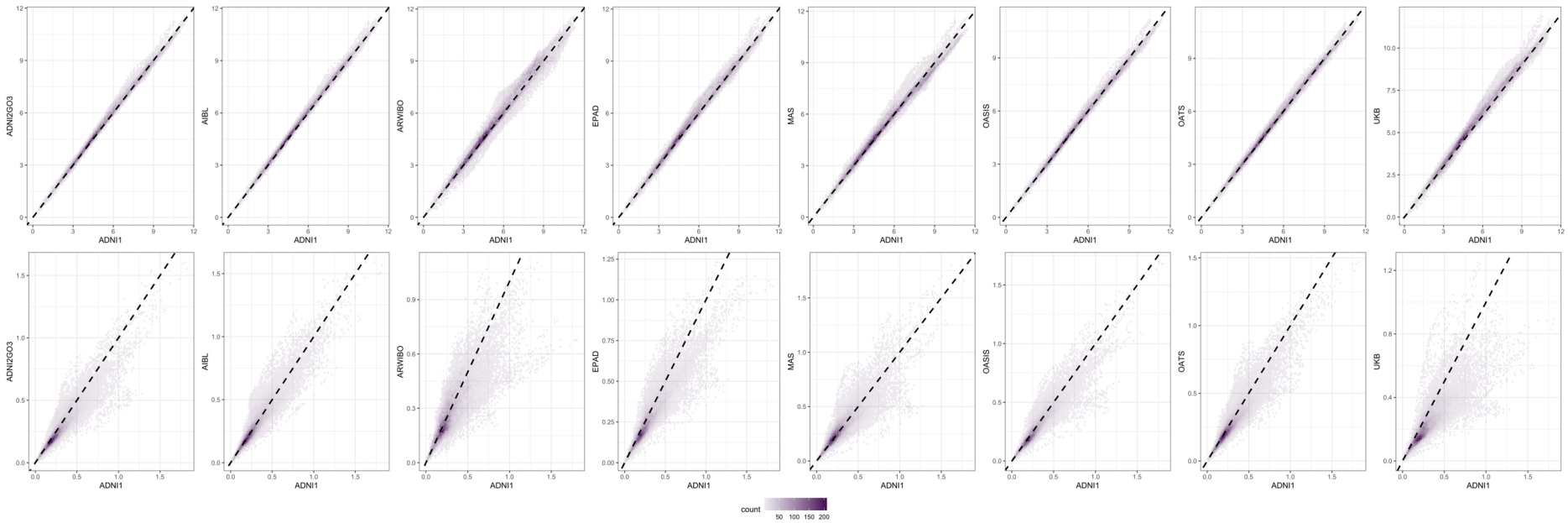
SFigure 2: Concordance of means (top panels) and variances (bottom panels) of the grey-matter vertices of cortical thickness

The different panels correspond to the different cohorts. ADNI1 is used as the reference (x-axis). The dotted line represents the identity.

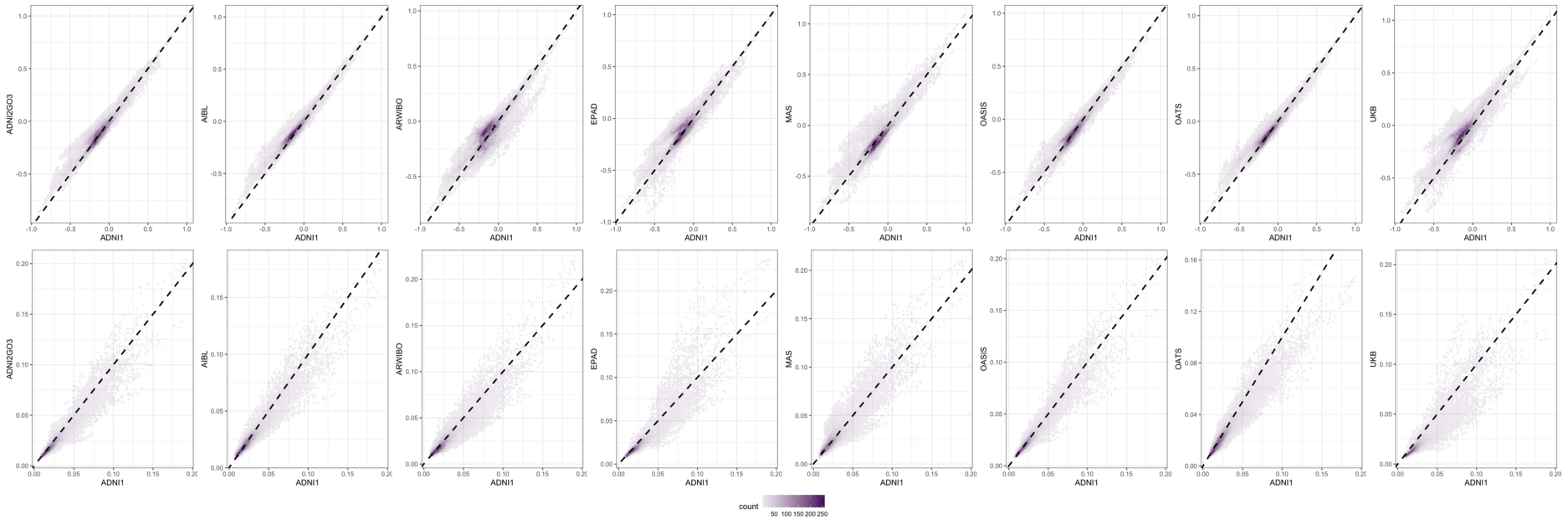


SFigure 3: Concordance of means (top panels) and variances (bottom panels) of the grey-matter vertices of cortical surface area

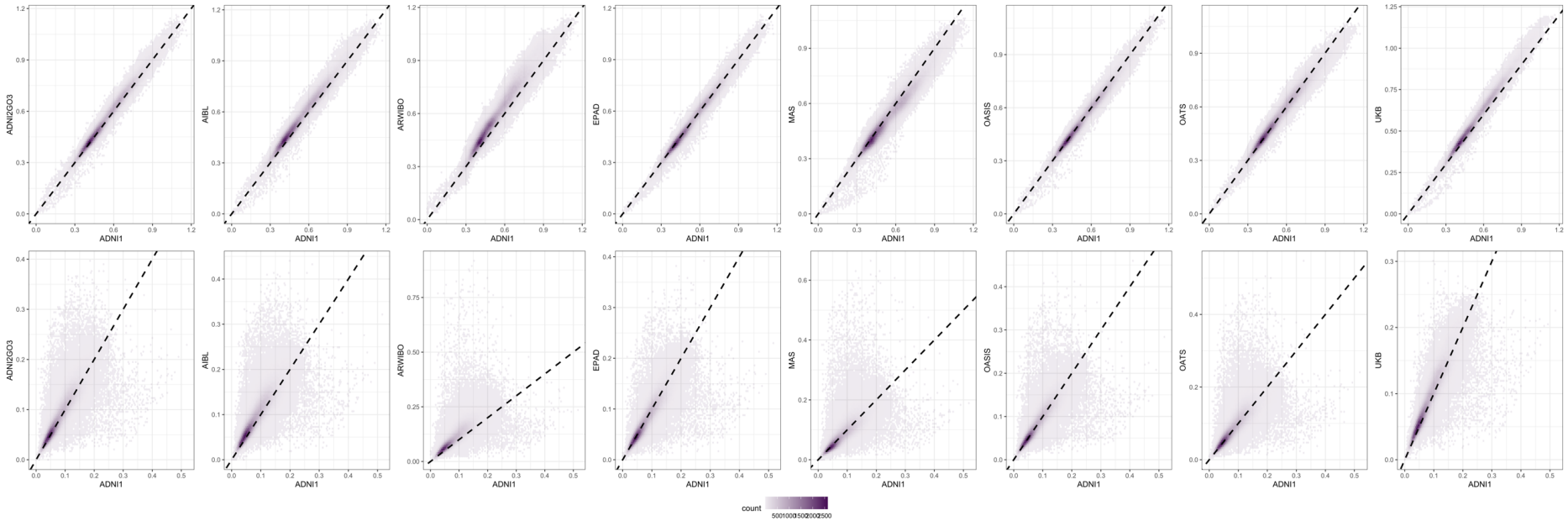
The different panels correspond to the different cohorts. ADNI1 is used as the reference (x-axis). The dotted line represents the identity.



SFigure 4: Concordance of means (top panels) and variances (bottom panels) of the grey-matter vertices of subcortical thickness.
The different panels correspond to the different cohorts. ADNI1 is used as the reference (x-axis). The dotted line represents the identity

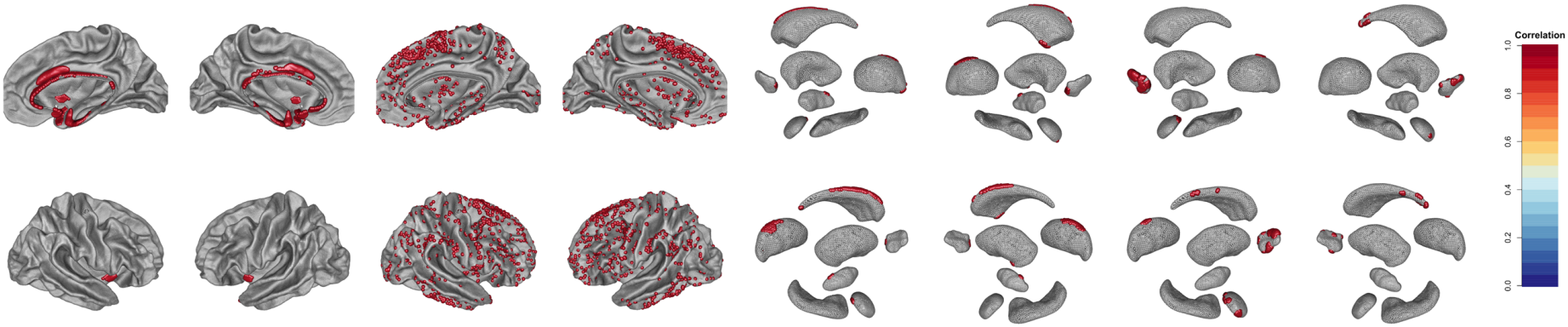


SFigure 5: Concordance of means (top panels) and variances (bottom panels) of the grey-matter vertices of subcortical surface area.
The different panels correspond to the different cohorts. ADNI1 is used as the reference (x-axis). The dotted line represents the identity .



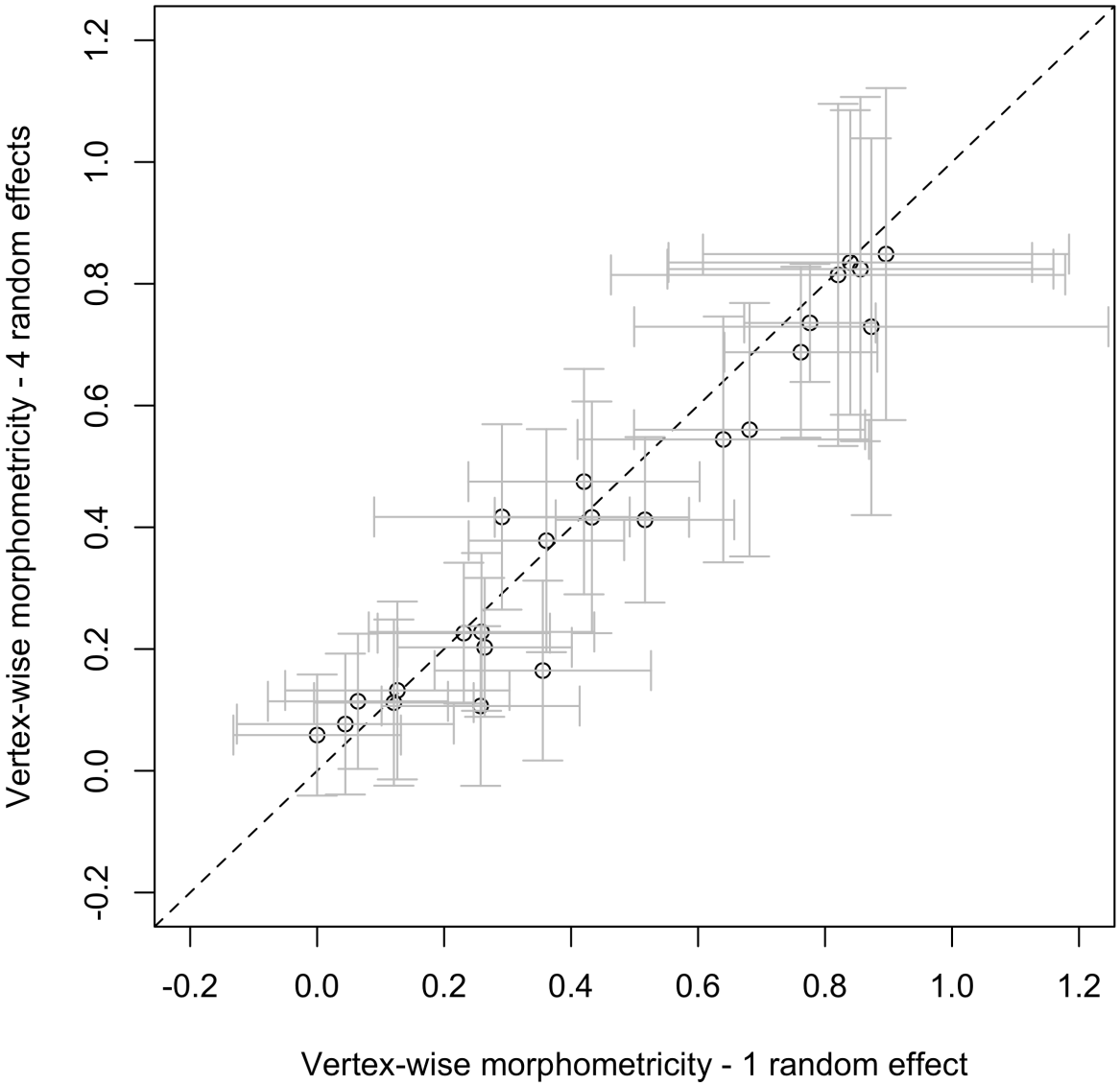
SFigure 6: Concordance of means (top panels) and variances (bottom panels) of the grey-matter vertices of cortical surface area – after removing "noisy vertices".

The different panels correspond to the different cohorts. ADNI1 is used as the reference (x-axis). The dotted line represents the identity. The figure shows the good concordance of mean and variances in cortical surface area when excluding the 1% of noisy vertices (see **SFigure 2** without exclusion).

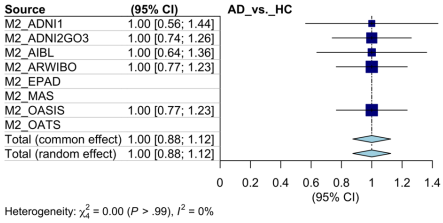


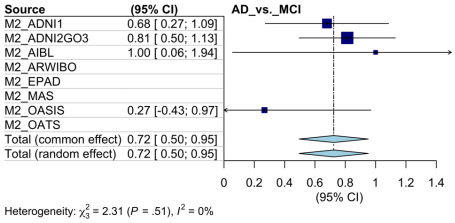
SFigure 7: Position of the noisy vertices for cortical thickness (L, R), cortical surface area (L, R), subcortical thickness (L, R) and subcortical surface area (L, R).

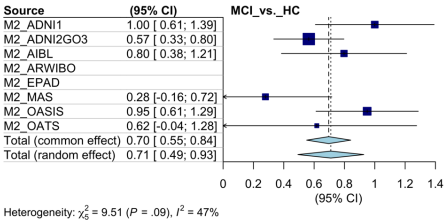
Top row shows the outside view of the cortex and subcortical structures; bottom row shows the inside view. Noisy vertices are shown in red. We labelled as "noisy" vertices that showed inconsistent variance from one sample to another. We only included healthy individuals to estimate the vertex means and variance, as different samples included different proportion of cases, for which we can expect brain atrophy, hence an impact on the mean and variances. For each pair of samples, we calculated the difference in vertex variance, and flagged the extreme vertices of the distribution (>6SD from mean). In practice this flags the vertices that are the most distant from the regression lines shown in **SFigure 1-4**. This led to label as noisy, 0.8% of the cortical surface vertices, 1.2% of the cortical thickness ones, 1.9% of the subcortical surface area and 1.4% of the subcortical thickness measurements.

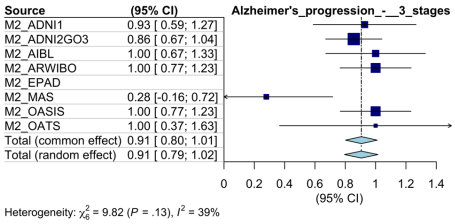


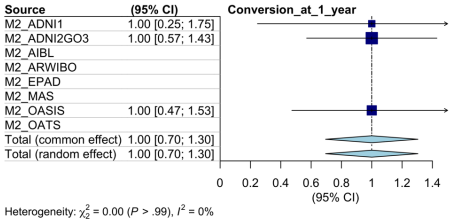
SFigure 8: Vertex-wise morphometricity estimated using two different mixed models. Each point corresponds to one of the 24 traits studied. On the x-axis, we show the morphometricity (and 95% confidence intervals in grey) estimated using a mixed model with a single random effect. The y-axis shows the morphometricity (and 95% confidence intervals in grey) from the model where each modality (cortical thickness, cortical surface area, subcortical thickness, subcortical surface area) is fitted as a specific random effect. The dashed line represents the identity. The morphometricity estimates have been meta-analysed across all the clinical samples.

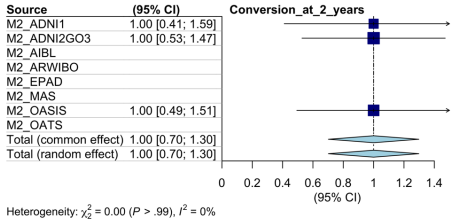


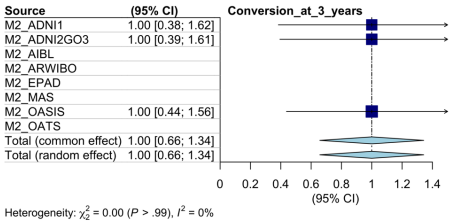


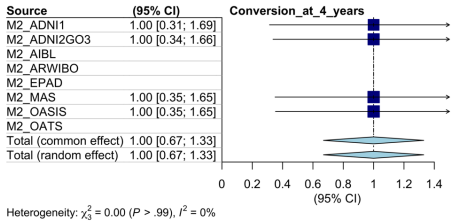


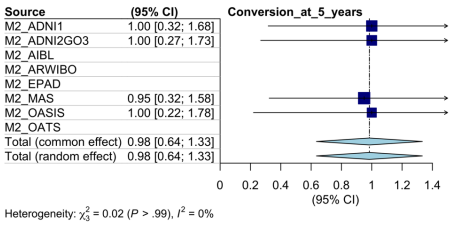


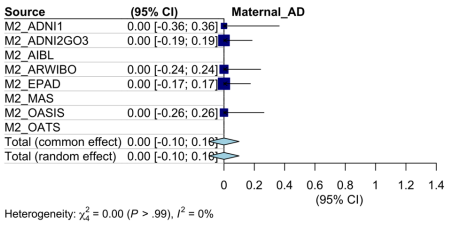


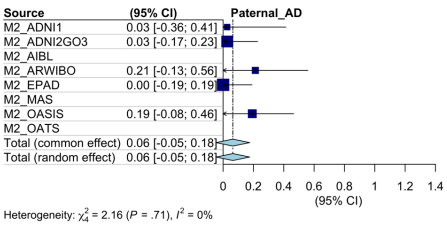






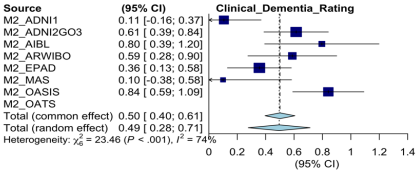


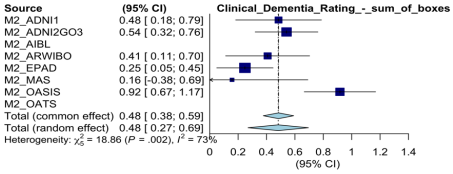


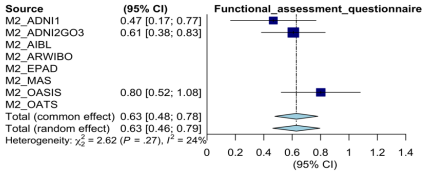


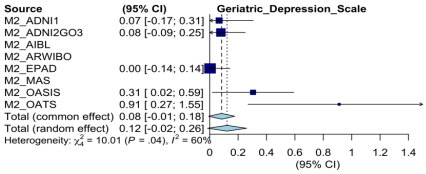
SFigure 9: Forest plots for the meta-analysis of vertex-wise morphometricity – disease status, conversion and family history

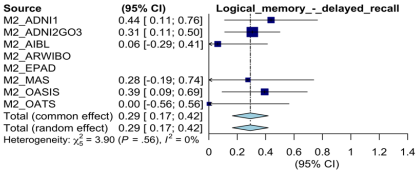
For each trait of interest, the panel shows the vertex-wise morphometricity estimated within each sample, the meta-analysed results and the level of between-study heterogeneity (estimated from the random effect meta-analysis model). Here, morphometricity was estimated with all vertices fitted in a single random effect), covariates include age, sex, site and global brain measurements.

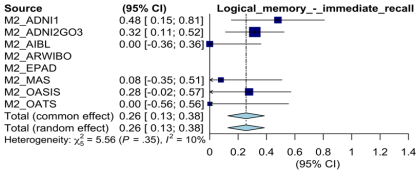


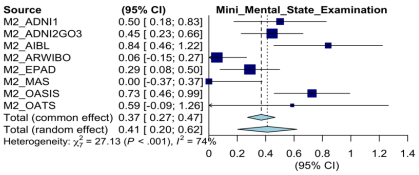


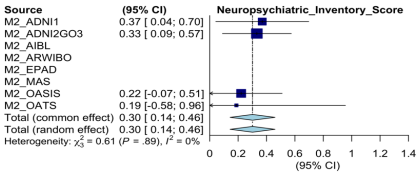


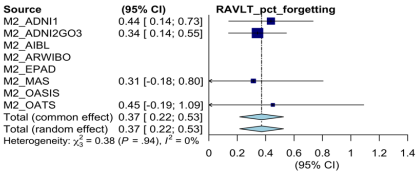


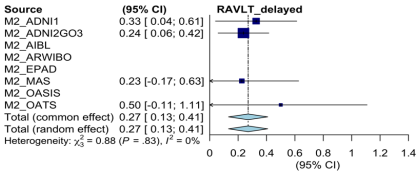


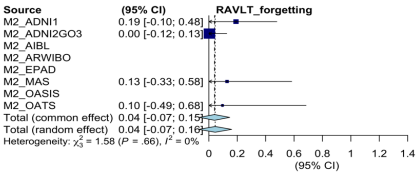


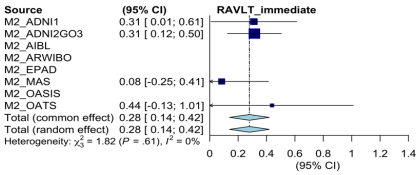


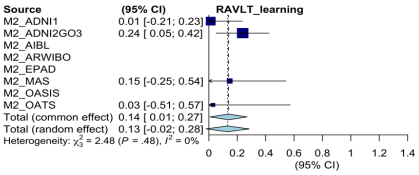








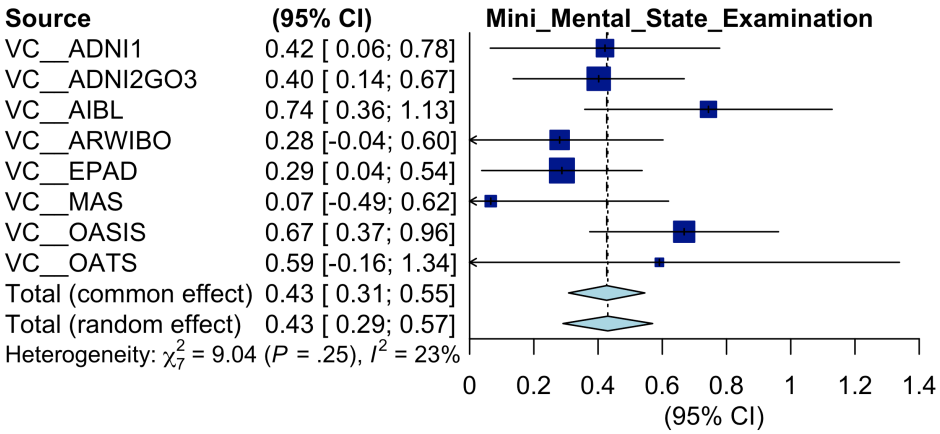


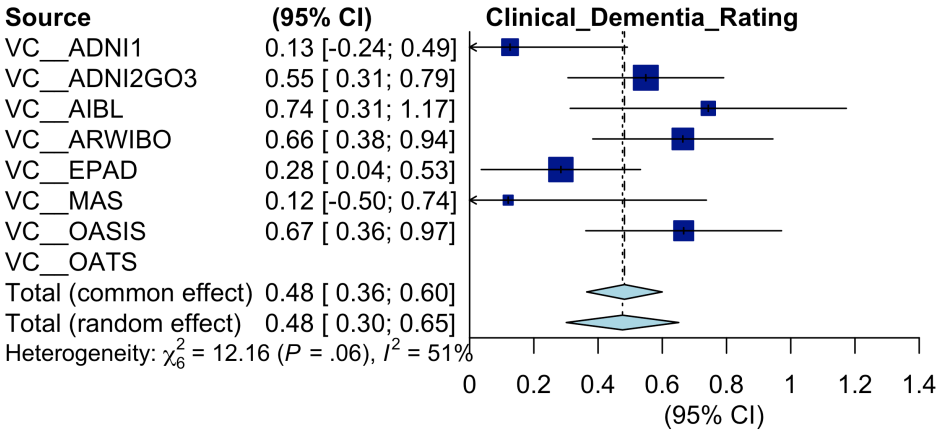


SFigure 10: Forest plots for the meta-analysis of vertex-wise morphometricity – neuropsychological scales

For each trait of interest, the panel shows the vertex-wise morphometricity estimated within each sample, the meta-analysed results, and the level of between-study heterogeneity (estimated from the random effect meta-analysis model). Here,

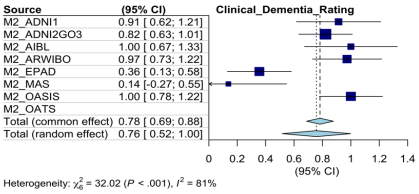
morphometricity was estimated with all vertices fitted in a single random effect), covariates include age, sex, site, global brain measurements and disease status to account for the fact the samples do not include the same proportion of Alzheimer's cases.



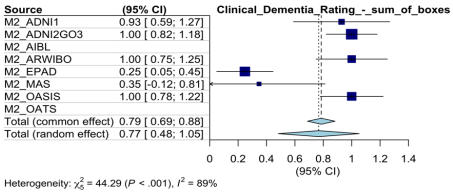


SFigure 11: Forest plots of vertex-wise morphometricity using a mixed models with 4 variance components

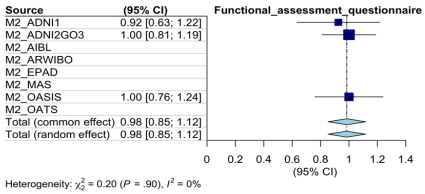
Compared to SFigure 9 (morphometricity estimated using a single random effect), the between-sample heterogeneity of morphometricity estimates is greatly reduced.



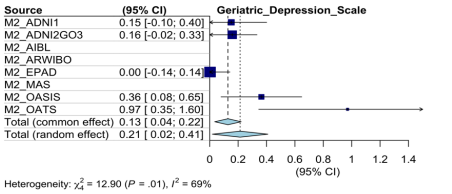
Heterogeneity: χ^2_6 = 32.02 (P < .001), I^2 = 81%



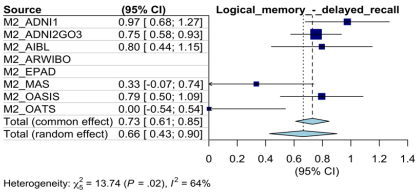
Heterogeneity: χ^2_5 = 44.29 (P < .001), I^2 = 89%



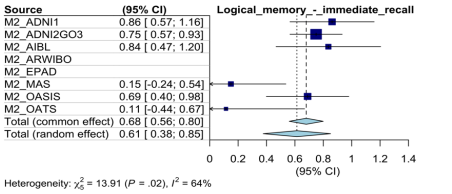
Heterogeneity: χ^2_2 = 0.20 (P = .90), I^2 = 0%



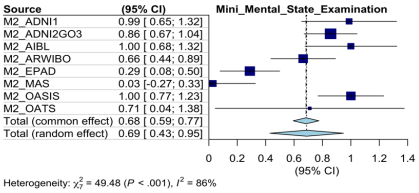
Heterogeneity: χ^2_4 = 12.90 (P = .01), I^2 = 69%



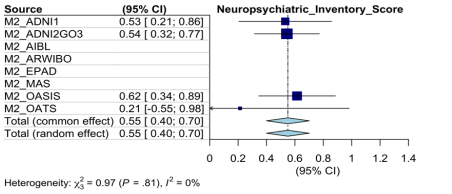
Heterogeneity: χ^2_5 = 13.74 (P = .02), I^2 = 64%



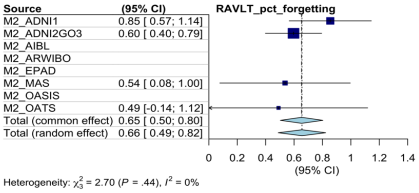
Heterogeneity: χ^2_5 = 13.91 (P = .02), I^2 = 64%



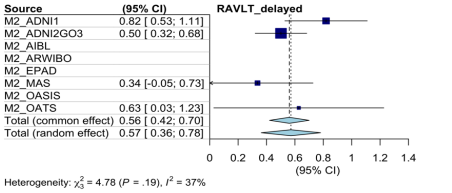
Heterogeneity: χ^2_7 = 49.48 (P < .001), I^2 = 86%



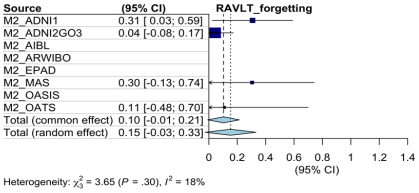
Heterogeneity: χ^2_3 = 0.97 (P = .81), I^2 = 0%



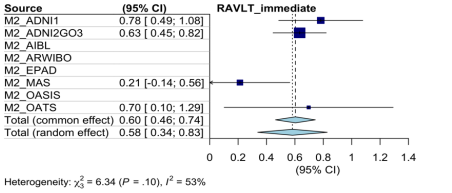
Heterogeneity: χ^2_3 = 2.70 (P = .44), I^2 = 0%



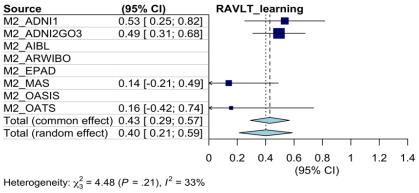
Heterogeneity: χ^2_3 = 4.78 (P = .19), I^2 = 37%



Heterogeneity: χ^2_3 = 3.65 (P = .30), I^2 = 18%



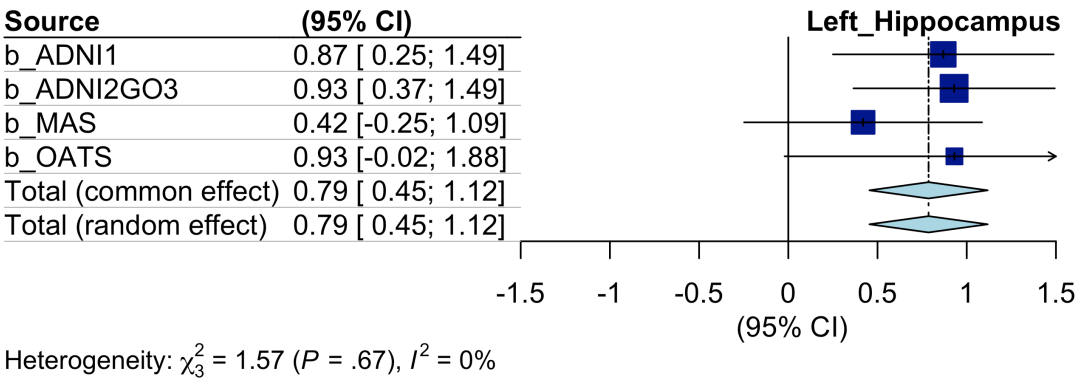
Heterogeneity: χ^2_3 = 6.34 (P = .10), I^2 = 53%



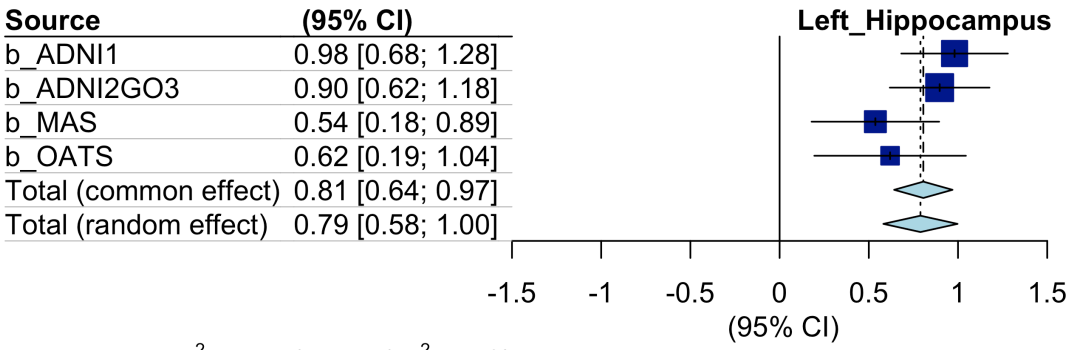
Heterogeneity: χ^2_3 = 4.48 (P = .21), I^2 = 33%

SFigure 12: Morphometricity estimates of neuropsychological scales without controlling for disease status

As expected, for the scores associated with Alzheimer's (i.e. most of the scores included), the morphometricity estimates are overestimated in cohorts that contained Alzheimer's cases (ADNI1, ADNI2Go3, AIBL, ARWIBO, OASIS) compared to the others (EPAD, MAS and OATS). This leads to a large between-study heterogeneity and an inflation of the overall meta-analysed morphometricity (See **SFigure 9** for results controlling for disease status).



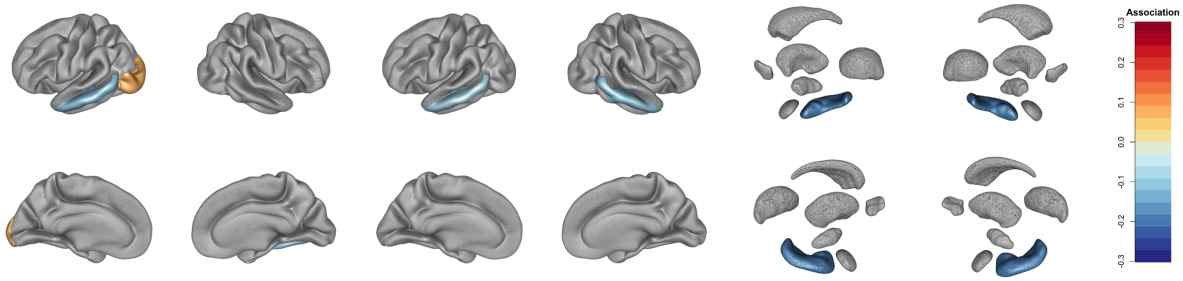
Heterogeneity: $\chi^2_3 = 1.57$ ($P = .67$), $I^2 = 0\%$



Heterogeneity: $\chi^2_3 = 4.65$ ($P = .20$), $I^2 = 35\%$

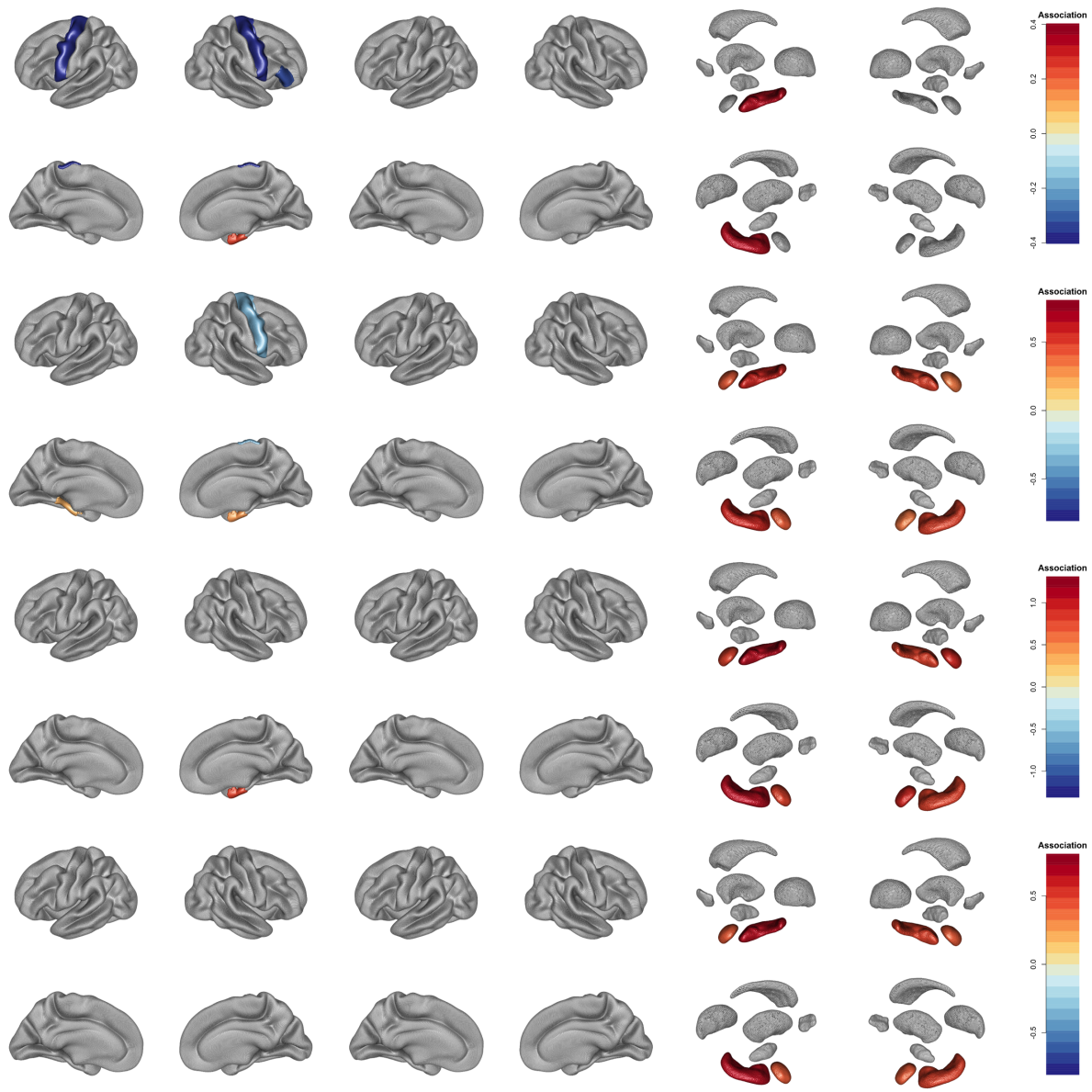
SFigure 13: Forest plots of the marginal (multivariate) and univariate association between left hippocampal volume and RAVLT delayed recall score

The marginal association (i.e. controlling for all ROI measurements and covariates) is shown at the top, followed by the univariate association, which only controls for the covariates.



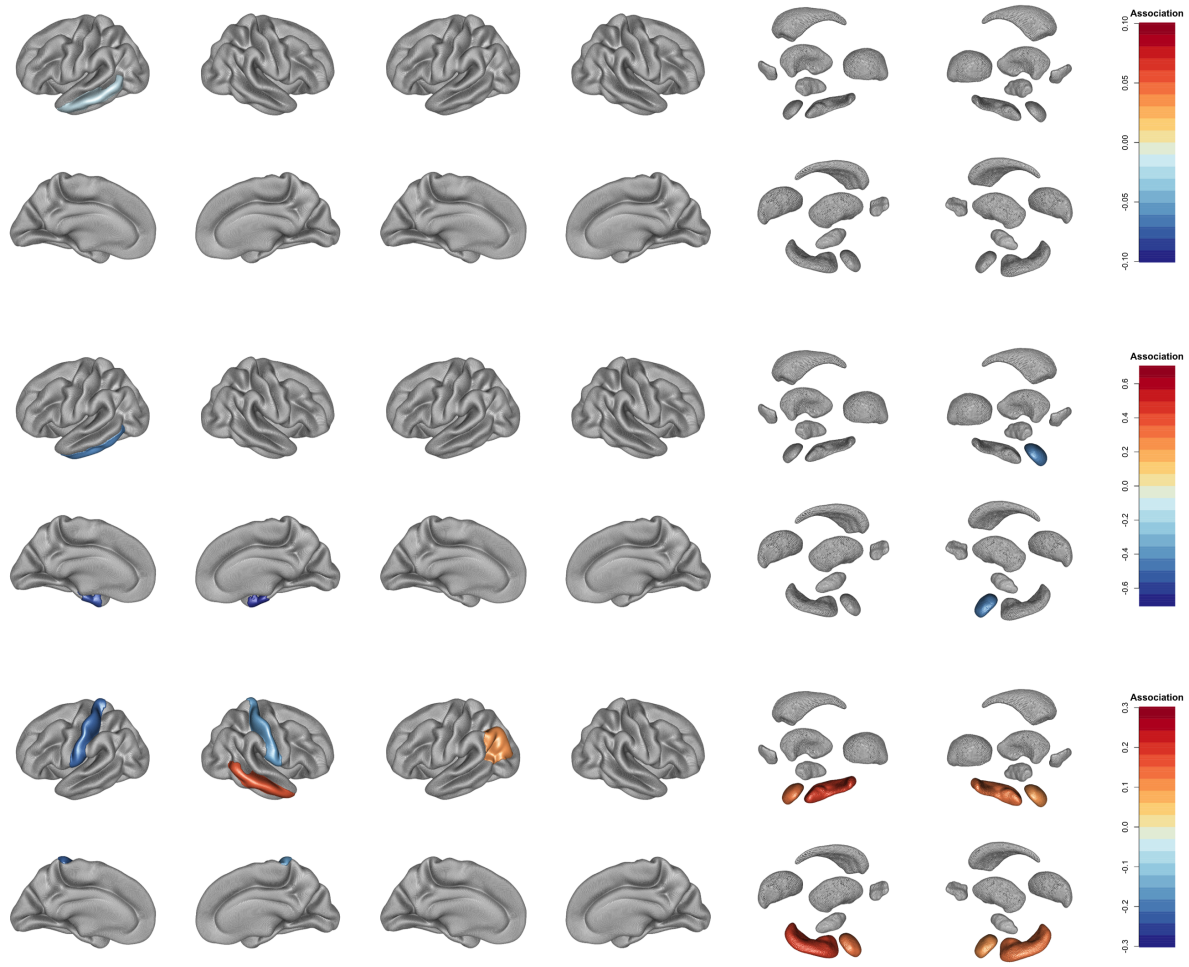
SFigure 14: Univariate ROI associations with Alzheimer's conversion within 3 years of imaging

Outside view (top panels) and Inside view (bottom panels). From left to right: left cortical thickness, right cortical thickness, left cortical surface, right cortical surface, left subcortical volumes and right subcortical volumes. We only show significant ROIs after multiple testing correction ($p<0.05/24/150$). The association betas correspond to the effect of 1 SD of ROI on the phenotype. Conversion at 1,2 and 4 years implicated similar ROIs (although not always all significant).



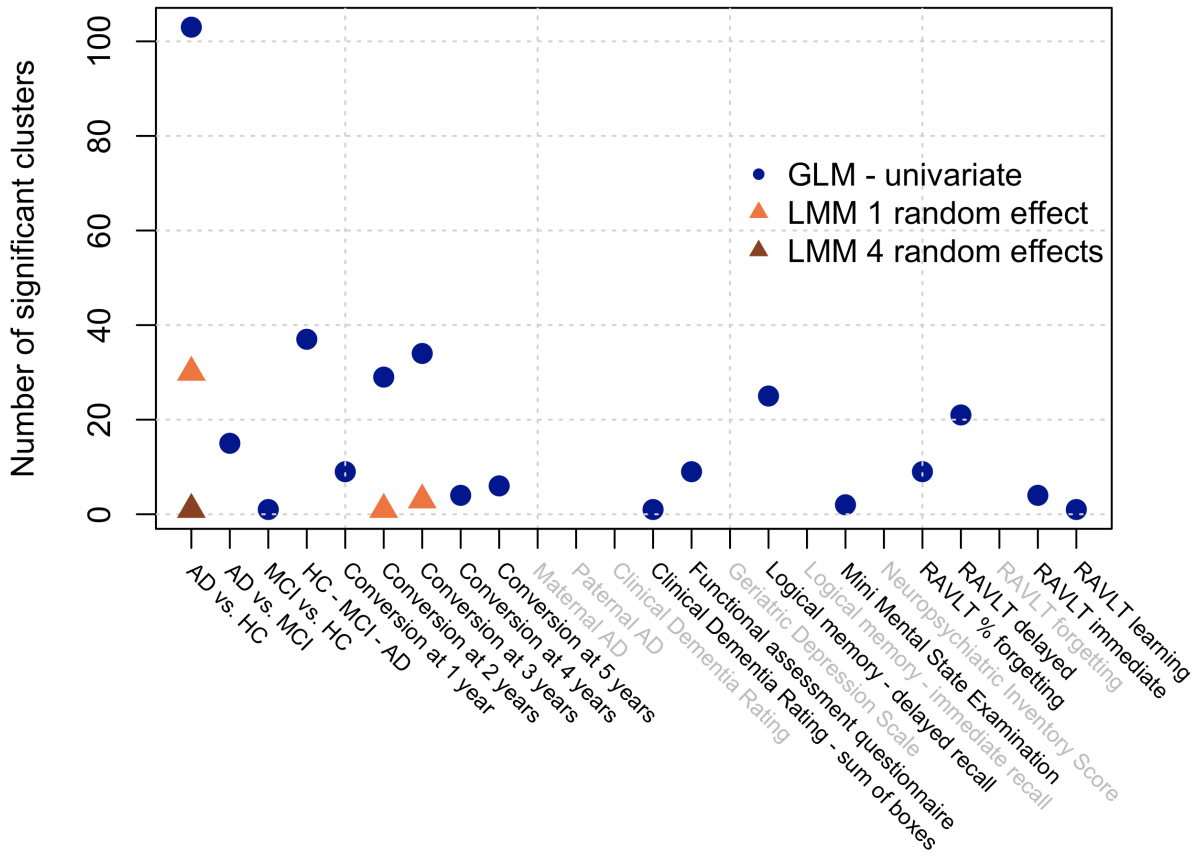
SFigure 15: Univariate ROI associations with memory scores

From top to bottom: Logical memory Immediate, Logical memory delayed, RAVLT score immediate, RAVLT score delayed. Outside view (top panels) and Inside view (bottom panels). From left to right: left cortical thickness, right cortical thickness, left cortical surface, right cortical surface, left subcortical volumes and right subcortical volumes. We only show significant ROIs after multiple testing correction ($p<0.05/24/150$). The association betas correspond to the effect of 1 SD of ROI on the phenotype. The colour scale may differ between phenotypes.



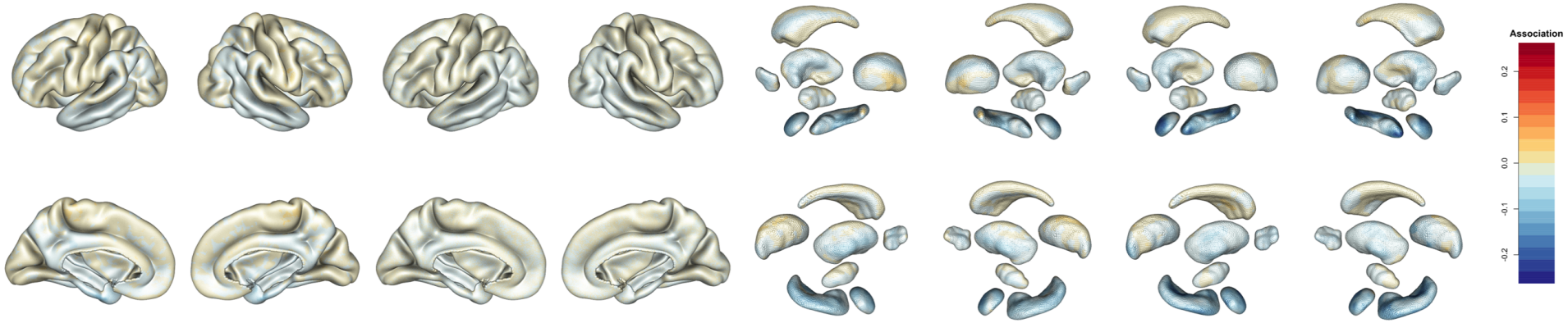
SFigure 16: Univariate ROI associations with neurological scales

From top to bottom: CDR, FAQ and MMSE.Outside view (top panels) and Inside view (bottom panels). From left to right: left cortical thickness, right cortical thickness, left cortical surface, right cortical surface, left subcortical volumes and right subcortical volumes. We only show significant ROIs after multiple testing correction ($p<0.05/24/150$). The association betas correspond to the effect of 1 SD of ROI on the phenotype. The colour scale may differ between phenotypes.

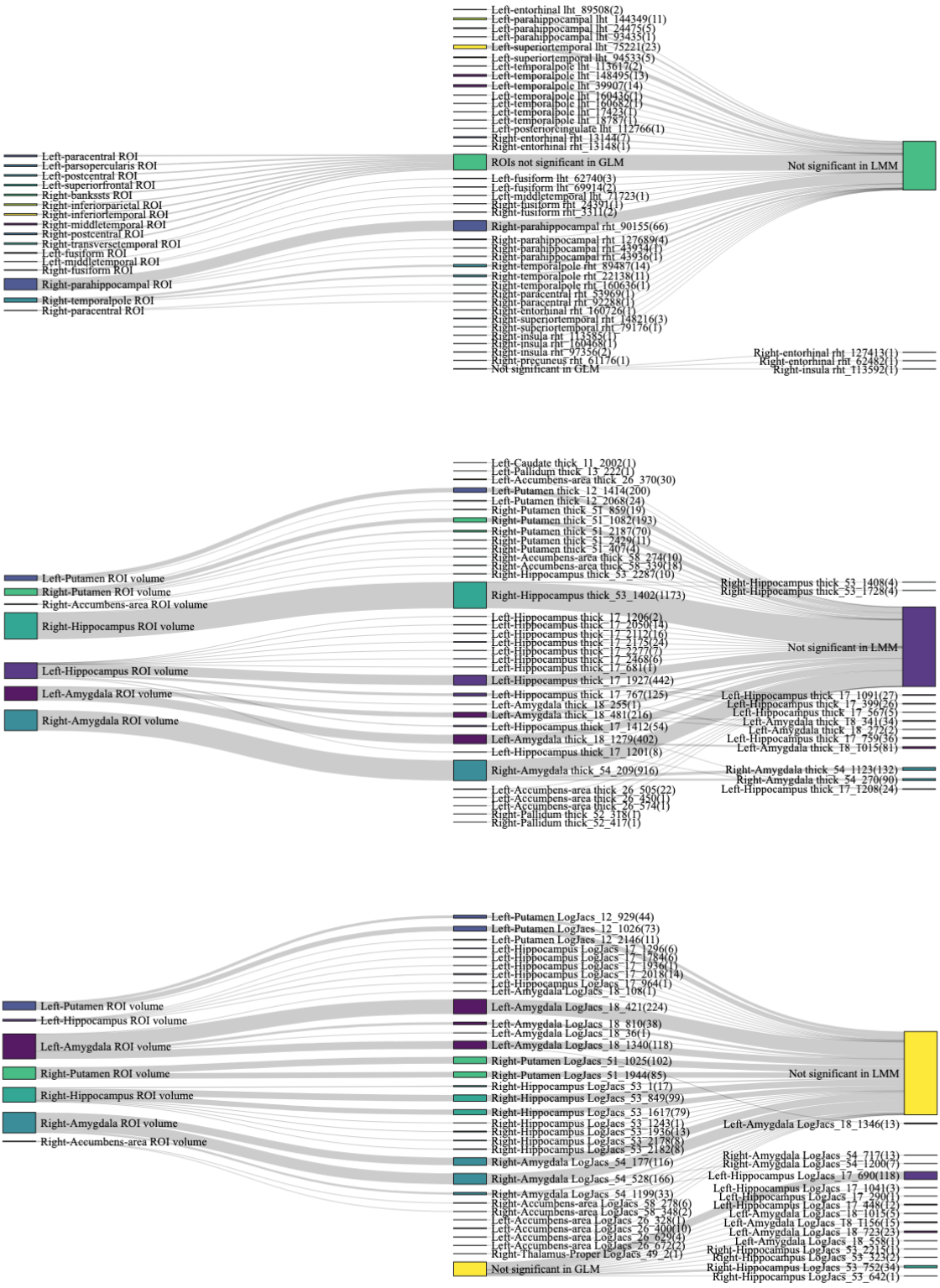


SFigure 17: Number of significant clusters (after Bonferroni correction) of the different models of vertex-wise associations

Traits for which no significant clusters were found using any of the methods are shown in grey.



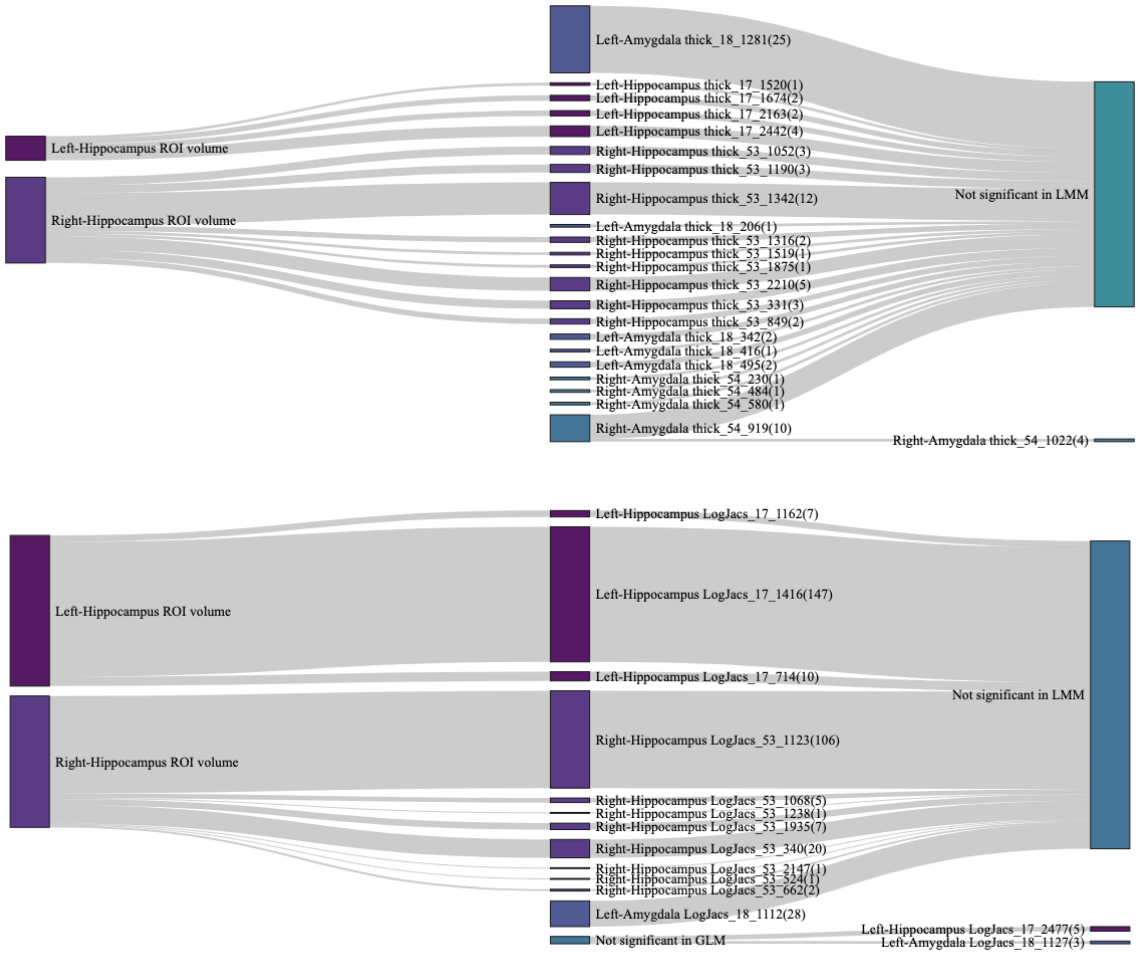
SFigure 18: Unthresholded map of association (betas) for AD vs. HC, using the GLM approach.



SFigure 19: Concordance of brain regions associated with Alzheimer's disease (AD vs. HC) across the different analyses (ROI and vertex-wise), by type of measurement.

From top to bottom: cortical thickness associations, subcortical thickness and subcortical surface area. Cortical surface area

is not shown as only ROIs were identified. The boxes on the left represent the significant ROIs (note: for subcortical structures only volume was available and tested). Boxes in the middle show the clusters identified in mass-univariate analysis (GLM) while the boxes on the right indicate the clusters significant in the multi-vertex approach (LMM). The size of the GLM and LMM bars are proportional to the number of vertices in the cluster. GLM and LMM clusters names indicate the brain region they mostly belong to, the top vertex and the number of vertices in cluster (in parentheses).

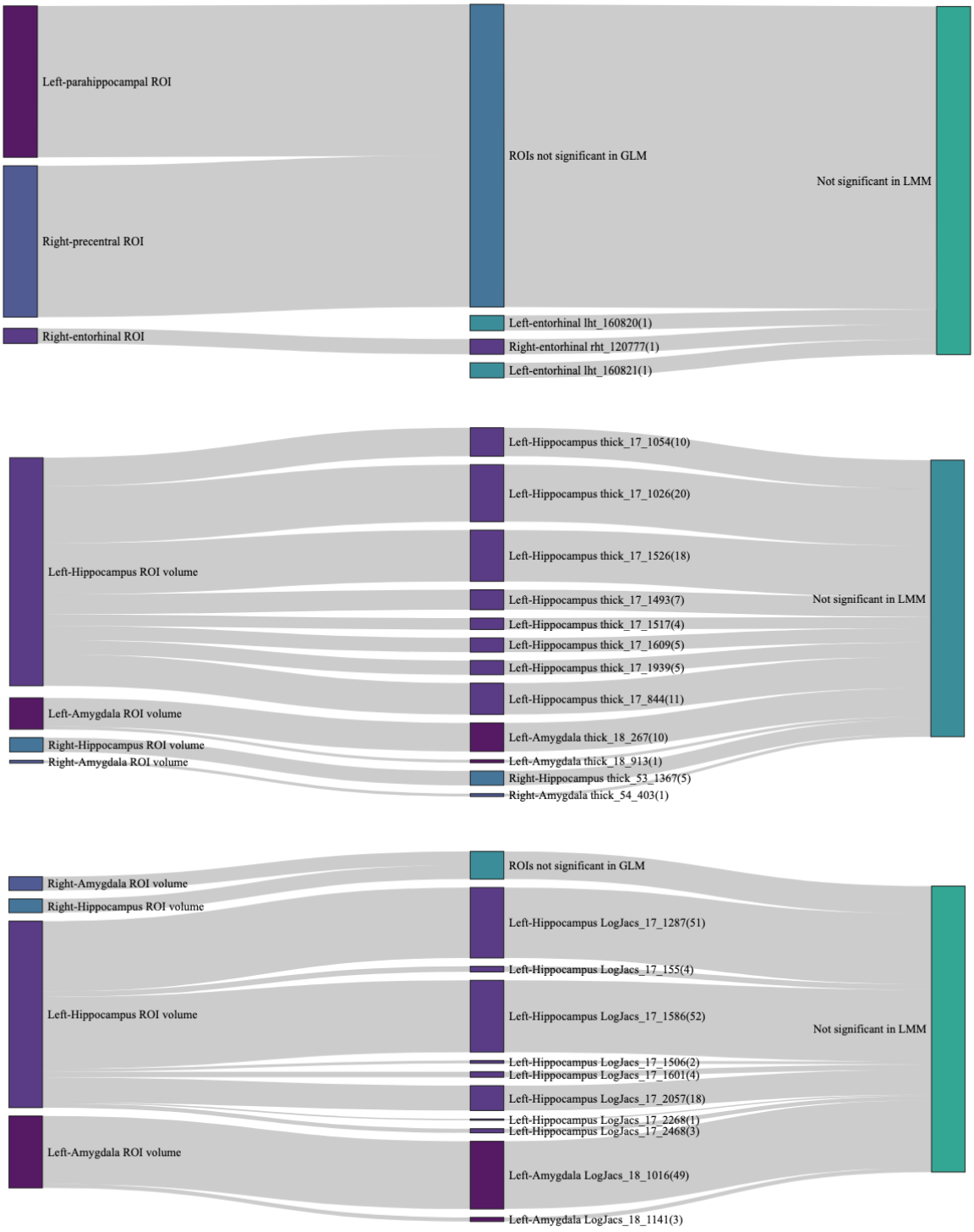


SFigure 20: Concordance of brain regions associated with Alzheimer's conversion (at 3 years) across the different analyses.

From top to bottom: subcortical thickness (thick) and subcortical surface area (LogJacs). The bars on the left represent the significant ROIs (note: for subcortical structures only volume was available and tested). Bars in the middle show the clusters identified in mass-univariate analysis (GLM) while the bars on the right indicate the clusters significant in the multivariate approach (LMM). The size of the GLM and LMM bars are proportional to the number of vertices in the cluster. GLM and LMM clusters names indicate the brain region they mostly belong to, the top vertex and the number of vertices in cluster (in parentheses).

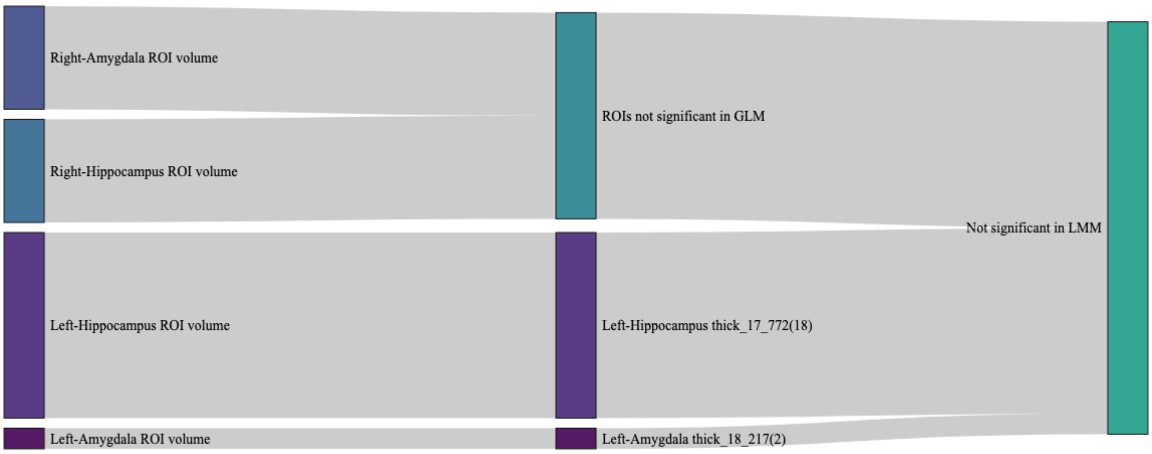
SFigure 21: Concordance of brain regions associated with Functional Activities Questionnaire (FAQ) score across the different analyses.

From top to bottom: cortical thickness, subcortical thickness (thick) and subcortical surface area (LogJacs). The bars on the left represent the significant ROIs (note: for subcortical structures only volume was available and tested). Bars in the middle show the clusters identified in mass-univariate analysis (GLM) while the bars on the right indicate the clusters significant in the multivariate approach (LMM). The size of the GLM and LMM bars are proportional to the number of vertices in the cluster. GLM and LMM clusters names indicate the brain region they mostly belong to, the top vertex and the number of vertices in cluster (in parentheses).



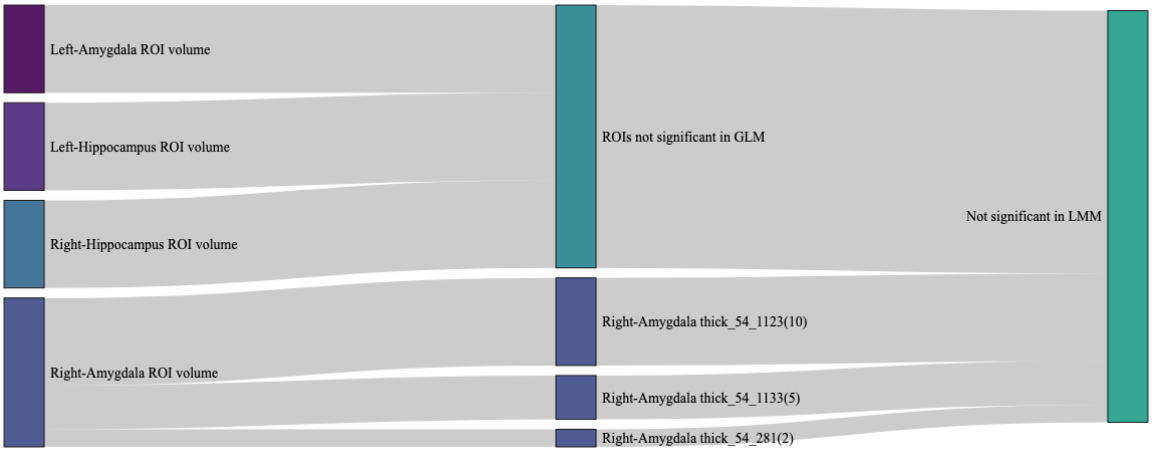
SFigure 22: Concordance of brain regions associated with Logical Memory (delayed recall) score across the different analyses.

From top to bottom: cortical thickness, subcortical thickness (thick) and subcortical surface area (LogJacs). The bars on the left represent the significant ROIs (note: for subcortical structures only volume was available and tested). Bars in the middle show the clusters identified in mass-univariate analysis (GLM) while the bars on the right indicate the clusters significant in the multivariate approach (LMM). The size of the GLM and LMM bars are proportional to the number of vertices in the cluster. GLM and LMM clusters names indicate the brain region they mostly belong to, the top vertex and the number of vertices in cluster (in parentheses).



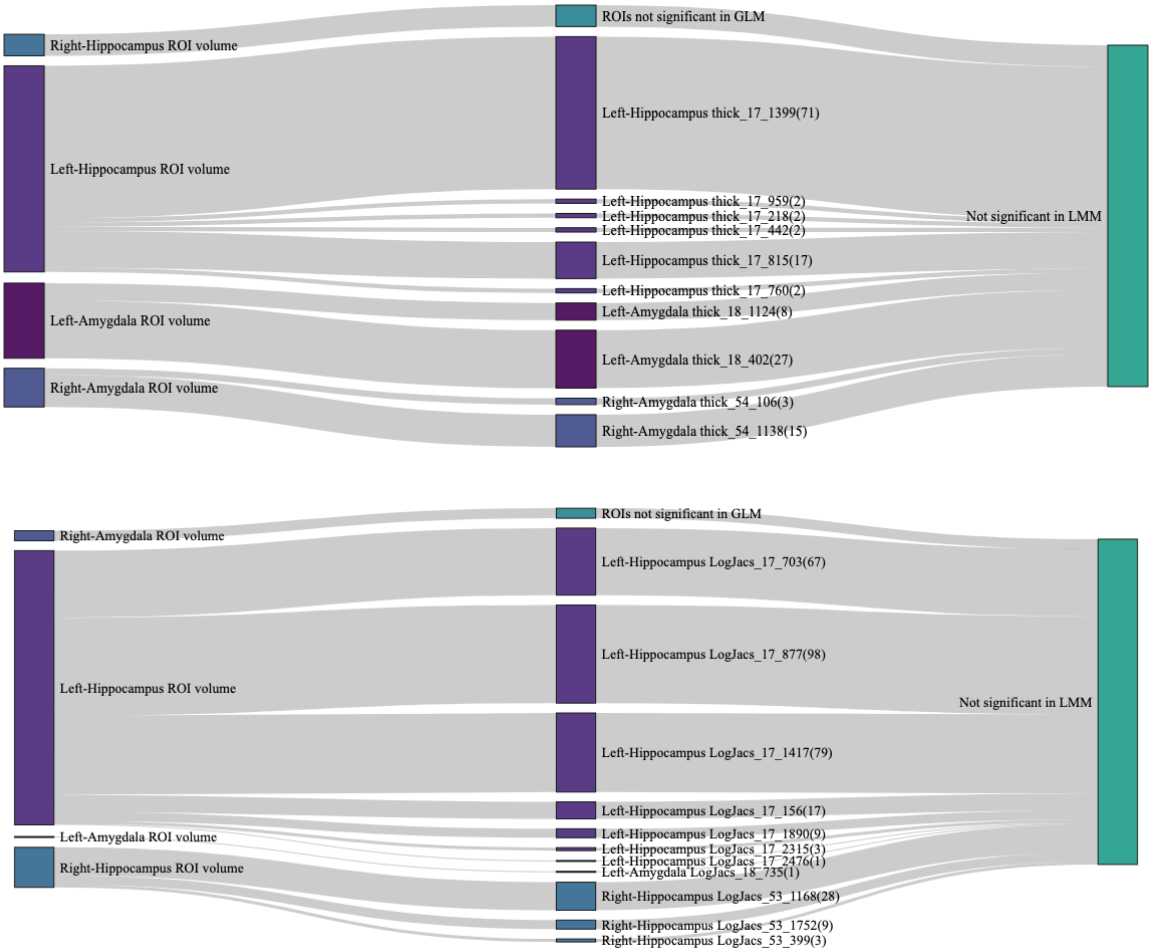
SFigure 23: Concordance of brain regions associated with MMSE score across the different analyses.

Subcortical thickness only. The bars on the left represent the significant ROIs (note: for subcortical structures only volume was available and tested). Bars in the middle show the clusters identified in mass-univariate analysis (GLM) while the bars on the right indicate the clusters significant in the multivariate approach (LMM). The size of the GLM and LMM bars are proportional to the number of vertices in the cluster. GLM and LMM clusters names indicate the brain region they mostly belong to, the top vertex and the number of vertices in cluster (in parentheses).



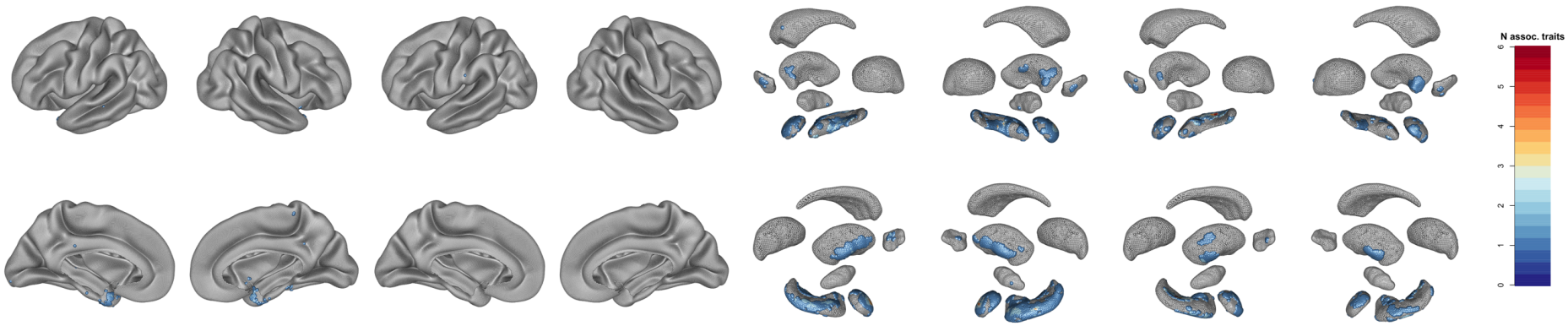
SFigure 24: Concordance of brain regions associated with RAVLT immediate memory score across the different analyses.

Subcortical thickness only. The bars on the left represent the significant ROIs (note: for subcortical structures only volume was available and tested). Bars in the middle show the clusters identified in mass-univariate analysis (GLM) while the bars on the right indicate the clusters significant in the multivariate approach (LMM). The size of the GLM and LMM bars are proportional to the number of vertices in the cluster. GLM and LMM clusters names indicate the brain region they mostly belong to, the top vertex and the number of vertices in cluster (in parentheses).



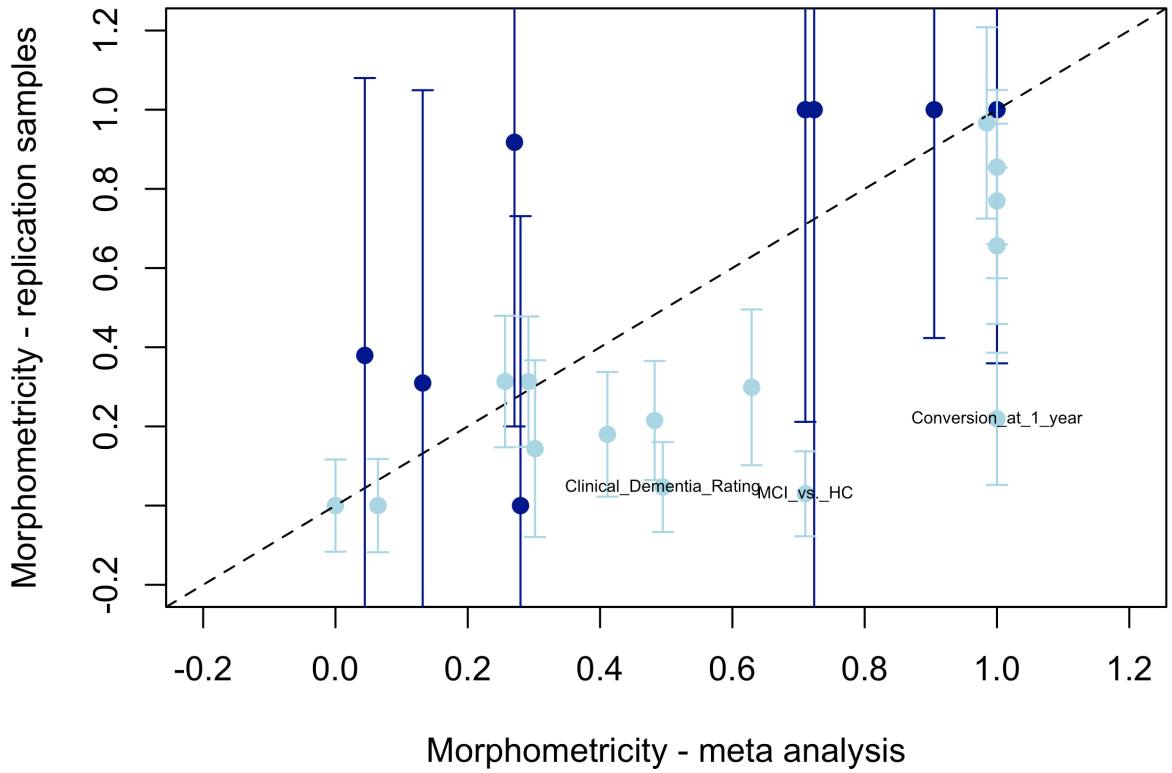
SFigure 25: Concordance of brain regions associated with RAVLT delayed memory score across the different analyses.

From top to bottom: subcortical thickness (thick) and subcortical surface area (LogJacs). The bars on the left represent the significant ROIs (note: for subcortical structures only volume was available and tested). Bars in the middle show the clusters identified in mass-univariate analysis (GLM) while the bars on the right indicate the clusters significant in the multivariate approach (LMM). The size of the GLM and LMM bars are proportional to the number of vertices in the cluster. GLM and LMM clusters names indicate the brain region they mostly belong to, the top vertex and the number of vertices in cluster (in parentheses).



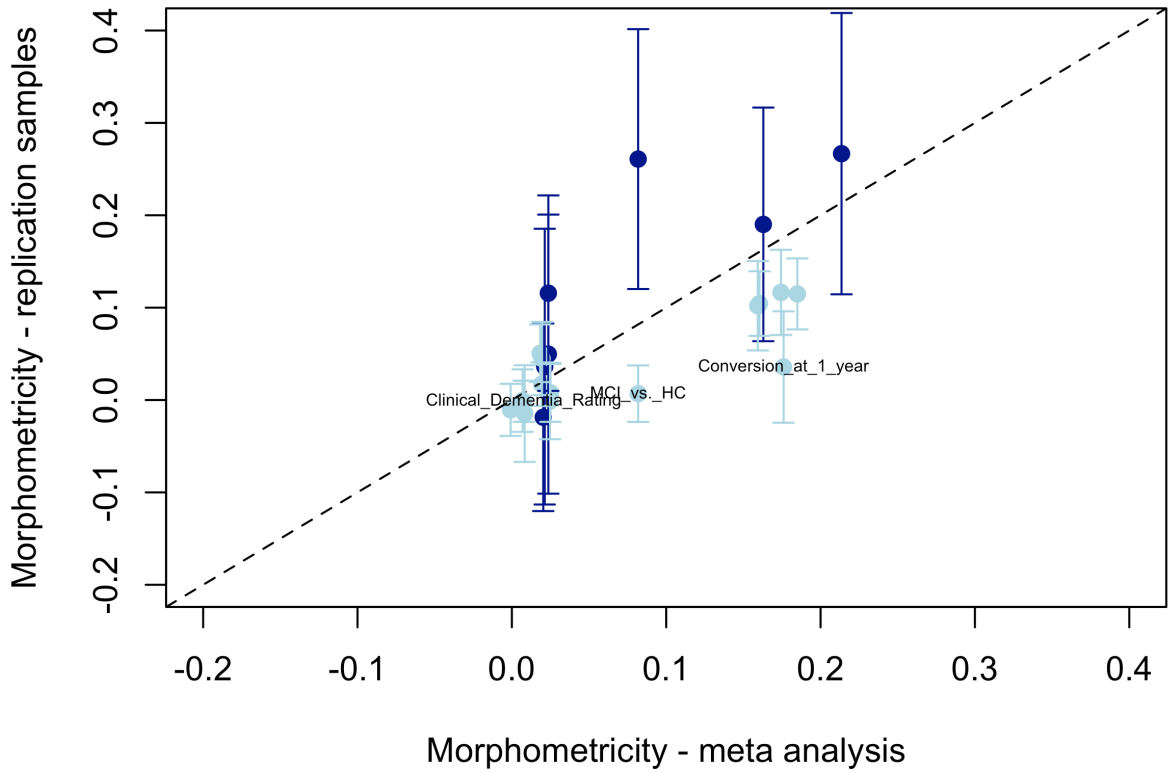
SFigure 26: All associated vertex-wise measurements in any of the 24 traits considered.

The colours show the number of traits associated with each vertex-wise measurement. Outside view (top panels) and Inside view (bottom panels). From left to right: left cortical thickness, right cortical thickness, left cortical surface, right cortical surface, left subcortical thickness, right subcortical thickness, left subcortical surface area, and right subcortical surface area. The maximum was found in the left hippocampus surface area with 6 traits associated with the same region.



SFigure 27: Comparison of vertex-based morphometricity estimates between the discovery and replication samples.

Estimates from the discovery samples (meta-analysis) are shown in the x-axis, and replication estimates are shown as the y-axis. Dark blue indicates replication in PISA, while light blue corresponds to MEMENTO. The vertical bars show the 95% confidence intervals in the replication samples. We have highlighted the three MEMENTO results showing the most discrepancy between discovery and replication.



SFigure 28: Comparison of ROI-based morphometricity estimates between the discovery and replication samples.

Estimates from the discovery samples (meta-analysis) are shown in the x-axis, and replication estimates are shown as the y-axis. Dark blue indicates replication in PISA, while light blue corresponds to MEMENTO. The vertical bars show the 95% confidence intervals in the replication samples. We have highlighted the same MEMENTO phenotypes as in SFigure 24, which also show a drop in association using the ROI.

Supplementary Tables (Captions)

STable 1: Description of the samples included in the analyses.
We defined as "healthy controls" individuals without a diagnosis of AD or MCI, but they may differ between cohorts according to the recruitment strategies and screening. For example, MEMENTO recruited in memory clinics, and all HC have subjective cognitive decline, while ARWIBO controls were screened for a range of neurodegenerative disorders. See Appendix A for more details.

STable 2: Proportion of noisy vertices within each ROI, for each type of measurement
We used the Desikan atlas (Desikan et al., 2006) to define the cortical ROIs. ROIs that have less than 1% of the noisy measurements have been excluded from the table.
Noisy vertices are over-represented in some brain regions, which depended on the type of measurement considered. For cortical thickness measurements, noisy vertices were mostly located in the caudal anterior cingulate and enthorinal and to a lower extend in the posterior cingulate, insula, temporal pole and rostralanterior cingulate. For cortical surface area, noisy vertices were rather scattered across cortical regions. For subcortical thickness, noisy vertices were mostly located on the top edge of the caudate and thalamus. For subcortical surface area, noisy vertices were found in the right accumbens. Of note, noisy vertices were distributed in a symmetrical fashion between the left and right hemisphere. These observations confirm that some brain regions are more prone to measurement errors.

STable 3: Association R2 between traits and global brain measurements
Global brain measurements include ICV, Left and Right average cortical thickness and Left and Right cortical surface area. Association R2 was estimated using a linear model in R, controlling for other covariates (age, sex, site). The R2 presented here has been meta-analysed across all clinical cohorts.

STable 4: ROI based morphometricity
Association R2 was estimated using a linear model in R, controlling for other covariates (age, sex, site and global brain measurements). For neuropsychological scales, we also controlled for the AD and MCI status. The R2 presented here has been meta-analysed across all clinical cohorts.

STable 5: vertex based morphometricity
Association R2 was estimated using a linear mixed model (single random effect) in OSCA, controlling for other covariates (age, sex, site and global brain measurements). For neuropsychological scales, we also controlled for the AD and MCI status. The R2 presented here has been meta-analysed across all clinical cohorts.

STable 6: Traits associations with global brain measurements
The association with global measurement was tested in a linear model, that controls for the other covariates (age, sex, site). All global measurements were fitted jointly in the model. For neuropsychological scales, we also controlled for the AD and MCI status.

STable 7: Significant trait-ROI associations in a univariate model
The association with each ROI measurement is tested in a separate model, that only controls for the covariates (age, sex, site and global brain measurements). For

neuropsychological scales, we also controlled for the AD and MCI status. Associations in the replication (PISA and MEMENTO) samples are also shown.

STable 8: Number of significant vertices and clusters, after Bonferroni correction, using the different vertex-wise association models

STable 9: Description of the significant clusters associated with the traits of interest using a GLM univariate analysis
The significant clusters are summarized by their top vertex (smallest pvalue) and the ROI they belong to. Associations in the replication (PISA and MEMENTO) samples are also shown.

STable 10: Description of the significant clusters associated with the traits of interest using a LMM multivariate analysis (single random effect)
The significant clusters are summarized by their top vertex (smallest pvalue) and the ROI they belong to. Associations in the replication (PISA and MEMENTO) samples are also shown.

STable 11: Vertex-wise measurements associated with 2 or more traits in the mass univariate (GLM) approach.
All associations are significant after controlling for multiple testing. Association effect sizes (b: effet on outcome trait for 1 SD of vertex-wise measurement) are shown and indicate the direction of effect.

STable 12: Same trait-prediction into PISA and MEMENTO. Linear predictors were constructed from the significant ROI or vertex-wise measurements in the meta-analysis of clinical samples. For the vertex-wise measurements, we selected the top vertex in each cluster (smallest pvalue). The effect size of each ROI or vertex-wise measurements corresponded to that estimated in the meta-analysis. Empty rows correspond to the case where no ROI or vertices reach significance. R2: prediction R2, controlling for all covariates, SE (SE_R2) and confidence intervals (CIL- lower bound, CIU upper bound) estimated by bootstrap, and pvalue from likelihood ratio test.

STable 13: Predictive ability of the AD vs. HC predictor in the non-diseased PISA and MEMENTO individuals.
Linear predictors of Alzheimer's disease were constructed from the significant ROI or vertex-wise measurements in the meta-analysis of clinical samples. For the vertex-wise measurements, we selected the top vertex in each cluster (smallest pvalue). The effect size of each ROI or vertex-wise measurements corresponded to that estimated in the meta-analysis. R2: prediction R^2, controlling for all covariates, SE (SE_R2) and confidence intervals (CIL- lower bound, CIU upper bound) estimated by bootstrap, and pvalue from likelihood ratio test. For PISA, a total of 33 traits and scores were tested, here we report the 9 traits for which at least one predictor reached significance ($p<0.05/33/3$). In MEMENTO we tested 36 traits and 14 reached significance ($p<0.05/36/3$).

Appendices

Appendix A. Sample description, including acquisition protocols of MRI images

ADNI1

The Alzheimer's Disease Neuroimaging Initiative, (ADNI, adni.loni.usc.edu) was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). See www.adni-info.org, for the latest information.

We considered the 819 participants imaged as part of the ADNI1 project and used their baseline brain MRI acquired using 1.5T scanners. MRI processing failed for 14 individuals, yielding a final sample of 805 participants from 58 north American sites. Our ADNI1 final sample comprised 476 female participants (58%), mean age was 75.2 (SD=6.8). The sample consisted in 229 (28%) healthy controls, 401 (49%) individuals with mild cognitive impairment (MCI) and 188 Alzheimer's disease (AD) cases (23%). Females were slightly over represented in the MCI group (64% vs 52% in the control group). We derived the AD conversion variables from the longitudinal clinical assessments (**Table 1**). Many neuropsychological scales/batteries were also available including the MMSE, CDR and FAQ (**Table 1**).

The ADNI1 images were acquired on various 1.5T scanners, which depended on the contributing sites. See https://adni.loni.usc.edu/help-faqs/adni-documentation/ for a list of the scanner used. T1w images (MP-RAGE) were acquired for all participants, with full information available https://adni.loni.usc.edu/help-faqs/adni-documentation/ .

ADNI2+Go+3

We considered participants recruited as part of the later ADNI cohorts (ADNI GO, 2 and 3 [currently ongoing]), who were all imaged using 3T MRI scanners. Baseline MRI scans were available for 1,449 individuals. Our final sample, comprised 1,410 unique individuals with complete MRI processing and clinical information. Participants were 72.1 years old on average (SD=7.2) and about half were female (709, 50.2%). The sample broke down into 577 controls (41.0%), 619 MCI (4.9%) and 195 Alzheimer's cases (13.8%), though participants were followed for 5+ years, which provides information about AD conversion (**Table 1)**. Many neuropsychological or cognitive scores (incl. MMSE, CDR, FAQ) were available on most of the participants (**Table 1**).

ADNI2+GO+3 images were acquired in 50+ sites across the US, on a variety of 3T scanners https://adni.loni.usc.edu/help-faqs/adni-documentation/ . T1w images (accelerated sagittal MPRAGE) varied in term of acquisition parameters https://adni.loni.usc.edu/help-faqs/adni-documentation/ .

AIBL

Data was collected by the AIBL study group across two Australian sites. AIBL study methodology has been reported previously(Ellis et al., 2009). We accessed the AIBL subset available via the ADNI application, which consisted in clinical and MRI imaging of 616 unique individuals (69% from site 1, 31% from site 2). The earlier scans from site 2 (111 images) were acquired using a 1.5T machine, while all other scans came from a 3T scanner. MRI processing did not complete for 10 individuals, resulting in a final sample of 606 participants. Our final sample was 72.8 years old on average (SD=6.7) and comprised 55% of women (335). The sample broke down into 443 healthy controls, 89 MCI and 72 AD cases. Diagnosis was missing for two individuals. MMSE and CDR scores were available for all participants (**Table 1**). Information about AD conversion were limited due to limited longitudinal follow up (**Table 1**).

The AIBL MRI images were acquired using ADNI protocols. i.e. using a ADNI 3-dimensional (3D) Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence, with 1 × 1 mm in-plane resolution and 1.2 mm slice thickness, TR/TE/TI = 2300/2.98/900, flip angle 9°, and field of view 240 × 256 and 160 slices.

ARWIBO

ARWiBo (Alzheimer's Disease Repository Without Borders) gathers clinical, neuropsychological, EEG, neuroimaging, and biological data of patients with neurodegenerative diseases and healthy controls, collected over 10 years by a number of researchers of IRCCS Fatebenefratelli, Brescia, Italy(Frisoni et al., 2020; Riello et al., 2005). The overall goal of ARWiBo is to contribute, thorough synergy with neuGRID (https://neugrid2.eu), to global data sharing and analysis in order to develop effective therapies, prevention methods and a cure for Alzheimer' and other neurodegenerative diseases. The ARWIBO dataset comprised 934 individuals with complete T1w MRI processing. Mean age of the participants is 56.7 years (SD=16.0), and the majority (60.5%) are female. The sample comprises 130 AD cases (**Table 1**), but does not contain longitudinal follow-ups to asses AD conversion. Neuropsychological scales such as the MMSE, CDR or logical memory scores are available for most participants (**Table 1**). Finally, brain images have been acquired on 7 different machines/sites across the city of Brescia.

ARWIBO collected 3D MRI across 7 centres around Brescia, in Italy. Each centre used their own 1 or 1.5T scanner with the following acquisition parameters: 1) Brescia: Philips Gyroscan NT 1T, TR/TE/TI =203412/50009/NA, flip angle 30°, slice thickness 2.6mm; 2) Verona: Siemens Magnetom Impact 1T, TR/TE/TI =11.4/4.4/300, 8°, 1.33mm; 3) Brescia: Philips Intera 1T, TR/TE/TI =25/6892199993/0, 30°, 2.6mm; 4) Varese: Philips Eclipse 1.5T, TR/TE/TI =12/4/NA, 20°, 1mm; 5) Milan: Siemens Magnetom Vision 1.5T, TR/TE/TI =9.7/4/300, 12°, 1mm; 6) Brescia: GE signa HDxt 1.5T, TE/TR/TI=11584/5.06/600, 8°, 1mm; 7) Brescia: GE signa HDxt 1.5T, TE/TR/TI=8.92/4.2/26, 12°, 1mm.

EPAD

The European Prevention of Alzheimer's Dementia (EPAD) Longitudinal Cohort Study is a prospective multi-centric cohort, designed to progress disease modelling of preclinical and prodromal Alzheimer's disease(Solomon et al., 2019). We downloaded the latest tranche of the data (v1500), acquired across 20 EPAD partner sites across Europe(Lorenzini et al., 2021). A large variety of MRI scanners were used, from the three main manufacturers (GE, Philips and Siemens). Most of the scans (91%) were performed using a 3T machine and the remaining using a 1.5T(Lorenzini et al., 2021). Our final sample comprised 1,315 individuals, with complete processing of the MRI images collected at baseline. Participants were on average 66 years old (SD=6.6), with a majority of females (748 or 56.8%). All participants were deemed healthy controls, although seven converted to AD within one year and another two converted within two years of the baseline visit (**Table 1**). Several neuropsychological scores were available for the EPAD cohort, including the MMSE, CDR and GDS (**Table 1**).

The EPAD MRI images were acquired across 21 centres using a common scanning protocol. Thirteen sites used a Siemens machine (voxel size: 1.2 x 1.05 x 1.05, matrix size: 176 x 256 x 240, Sagittal acquisition, TE = 2.95, TR = 2300), seven sites used a Philips scanner (voxel size: 1.1 x 1.1 x 1.2, matrix size: 176 x 256 x 256, Sagittal acquisition, TE = 3.11, TR =1672.6/1526.6) and a single site used a GE machine (voxel size: 1.2 x 1.2 x 1.05, matrix size: 196 x 256 x 256, Sagittal acquisition, TE = 3.09, TR = 7.65.

MAS

The Sydney Memory and Ageing Study (MAS) is a prospective cohort established in 2005 that recruited 1,037 non-demented participants, although only 544 underwent a brain MRI at baseline(Sachdev et al., 2010; Tsang et al., 2013). Our final sample comprised 527 individuals with complete MRI processing, including 290 females (55%). Mean age of the participants was 78 (SD=4.7). The sample broke down into 175 MCI (33%) and 288 healthy controls (55%). Thirteen participants could not be classified due to missing neuropsychological data (missing or invalid IADL score, missing or invalid MMSE or score below 24)(Sachdev et al., 2010) Another 51 individuals were not classified (despite meeting the MCI criteria) because they reported being of non-English-speaking background, which may limit the validity of the assessment(Kochan et al., 2010). MAS participants were followed up every year which allowed us to measure AD conversion up to 7 years after the baseline scans(Tsang et al., 2013). MMSE and CDR scores were available for all participants (**Table 1**).

MRIs were collected on a Philips 3T Achieva Quasar Dual scanner located at the Prince of Wales Medical Research Institute, Sydney. The 3D T1-weighted structural (T1w TFE – turbo field echo) MRI were acquired coronally with repetition time TR=6.39 ms, echo time TE=2.9 ms, flip angle=8°, matrix size=256x256, field of view FOV=256x256x190 mm3, and slice thickness=1 mm with no gap between; yielding 1x1x1 mm3 isotropic voxels. Total scanning time (including all modalities: scout, T1w, T2w, DWI) was around 20 minutes.

OASIS3

The OASIS3 dataset comprises 1,098 participants, including 550 followed over several years, for a total of 2,168 data points(LaMontagne et al., 2018). First, we selected the 1,869 visits with complete MRI processing, and non-missing basic variables (AD status, age, sex). For each participant, we considered the first visit, resulting in 1,036 unique individuals. Our final sample was composed of 56% of women, mean age was 70.4 (SD=9.4). The sample broke done into 238 (22.9%) AD cases, 16 (0.015%) MCI, and 782 (75.5%) controls. Controls were significantly younger than cases (68.8 vs 75.5 years, p<1e-16). We used the longitudinal clinical assessments to derive variables of AD conversion at the different time windows. Finally, several neuropsychological scales were available to the analysis: MMSE, CDR, FAQ, GDS, NPIQ and UPRDS.

The OASIS3 dataset consisted in T1w MRI images, acquired on different Siemens scanners (1.5 and 3T). See https://www.oasis-brains.org/files/OASIS-3_Imaging_Data_Dictionary_v1.5.pdf for a list of the MRI machines and more information regarding the acquisition protocols.

OATS:

The Older Adults Twin Study (OATS) is a prospective twin sample that gathers participants from three Australian states (NSW, VIC and QLD), who were contacted thanks to the Australian Twin Registry (www.twins.org.au) (Sachdev et al., 2009, 2013). Four hundred participants (out of the 623 included) have been imaged using 1.5T MRI machines, resulting in 377 usable scans (accessible to download)(Koncz et al., 2018). After MRI processing, our final sample comprised 365 individuals, aged 70.3 (SD=5.1) on average, with a majority of females (65.8%). At baseline, only 6 individuals met the diagnostic criteria for AD, though 46 (12.7%) were labelled as MCI (**Table 1)**. Using the clinical follow-ups, we derived measures of AD conversion (2-5 years) and included neuropsychological scores such as the MMSE or GDR (**Table 1**).

Structural MRI, was acquired on 1.5 T scanners. Data collection relied on Siemens scanners with similar year of manufacture and upgrade for Victoria (Melbourne, Siemens Magnetom Avanto scanner) and Queensland (Brisbane, Siemens Sonata), while a Philips scanner (Gyroscan) was used in the third centre (New-South Wales: Sydney). The three centres used the same acquisition protocols, with standardisation of in-plane resolution and slice thickness(Koncz et al., 2018; Sachdev et al., 2009). A 3D phantom was used to detect variation across scanners (for correction of geometric distortion), and five volunteers were scanned on all three scanners for reliability measures. The standardized protocol is as follows: in-plane resolution 1x1 mm with slice thickness of 1.5 mm, contiguous slices, TR/TE/TI = 1530/3.24/780 ms, and flip angle = 8°(Koncz et al., 2018; Sachdev et al., 2009).

UK Biobank

The UKB imaging study comprised 40,016 individuals at the time of download (Nov. 2020), with data collected across three centres in Cheadle, Reading and Newcastle(Miller et al.,

2016). We selected individuals who were processed (by the UKB team) using both T1w and T2 FLAIR images, in order to ensure homogeneous processed data(Lindroth et al., 2019). This reduced our final sample to 37,644 individuals, with complete brain imaging/processing.

We constructed Alzheimer's disease proxy phenotypes similar to those used in a recent GWAS, which showed a high genetic correlation between the proxy phenotypes and AD case-control status (maternal AD: rg=0.91 (SE 0.24), paternal AD: rg=0.67 (0.40), both not significantly different from unity). We constructed maternal and paternal history of Alzheimer/dementia using the self-reported fields "Has/did your mother/father ever suffer from Alzheimer's disease/Dementia" (fields ID 20110 and 20107), and we aggregated the information collected in the different visits (assessment visit, follow-up, first imaging visit, and repeated imaging visit). We set the proxy phenotypes to missing, if the parent was under 60 years old (fields ID 1845 and 2946), had died before 60 (fields ID 3526 and 1807) or when their age was not answered.

Our final sample comprised 20,056 females (53%), and participants mean age was 63 years (SD=7.5) at the time of brain MRI. Individuals were mostly imaged in Cheadle (61%), followed by Newcastle (26%) and Reading (13%). A total of 6,712 (17.8%) individuals reported a maternal history of AD, while information was missing for 3,297 (8.8%) due to incomplete report or parental age lower than 60. In addition, 3,590 individuals (9.5%) reported a paternal history of AD, while the information was missing for 5,681 (15.1%). Fields 130836 ("Date Date F00 first reported (dementia in Alzheimer's disease)") and 42020 ("Date of Alzheimer's disease report") have been constructed using registers, primary care and hospital records, and may serve to evaluate AD conversion of the UKB participants. We found that only 6 individuals got a hospital diagnosis of AD up to 6 years after the MRI (**Table 1**).

The UKB images were acquired using SIEMENS MAGNETOM Skyra (syngo MR D13) scanners, present in each assessment centre. The T2 FLAIR images (3D SPACE, sagittal) were acquired with a voxel resolution of 1.05x1x1 mm, FOV: 192x256x256 matrix, duration: 6 minutes, in-plane acceleration iPAT=2, partial Fourier = 7/8, fat saturation, elliptical k-space scanning, prescan-normalise, TR=2000, N=208, TI=880ms. The T1w images (3D MPRAGE, sagittal) had a resolution of 1x1x1 mm; FOV: 208x256x256 matrix, scan duration: 5 minutes, in-plane acceleration iPAT=2, prescan-normalise; TR=5000, TE=395, N=192, TI=1800ms.
See https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=20253 and
https://www.fmrib.ox.ac.uk/ukbiobank/protocol/index.html

MEMENTO

Memento is a French cohort that recruited individuals with subjective cognitive complaints or mild cognitive impairment from nationwide memory clinics(Dufouil et al., 2017). A total of 2,323 individuals were recruited, of which 1,880 included an exploitable MRI T1w image. This final sample included 63% of females and participants were on average 70 years old (SD=8.7). Due to recruitment, 1,575 (83%) of the individuals were classified as MCI at baseline, and 305 (16%) were controls (although with subjective complaints). Participants were followed for more than 5 years, and 201 received a diagnosis of Alzheimer's within this time frame (**Table**

1). The MMSE, CDR were collected on almost all participants, as well as many other neuropsychological scales. Finally, family history showed that 6.8% (128) of the participants had a father diagnosed with Alzheimer's, and 19.6% (365) had a mother with Alzheimer's.

MEMENTO consisted of a multicentric MRI collection of 3D T1-weighted acquired using a 9:00 minutes acquisition protocol. Images were acquired on a range of 3T or 1.5T machines, which depended on the centre. T1w images were acquired in the sagittal plane, with 1 mm slices for 3T scanners and 1.3 mm slices for 1.5T scanners. The resolution was isotropic in the acquisition plane (sagittal). Acceleration or averaging was not used. See (Régy et al., 2022) for all details.

PISA

PISA is the **"**Prospective Imaging Study of Ageing: Genes, Brain and Behaviour", an Australian prospective cohort that follows individuals with high and low genetic risk of dementia (based on APOE status and polygenic risk score), as well as a subset of Alzheimer's cases and MCIs(Lupton et al., 2020). PISA recruited 3,800 individuals, although only a subset of 299 has currently undergone MRI imaging. The final sample, with complete demographics and MRI processing comprised 267 individuals, which includes 193 females (72.3%), 26 Alzheimer's cases (9.7%), 21 MCI (7.9%) and 220 healthy controls (82.4%). Mean age of the participants was 60.6 (SD=6.9). In addition, PISA collected extensive neuropsychological scores mapping most dimensions of cognition, including the RAVLT (Lupton et al., 2020).

PISA collected 3D MP2RAGE images (BW = 240 Hz/Px, 3xGRAPPA acceleration) over a 9:02min sequence with FOV: 256 × 240 × 192 (voxel size 1 × 1 × 1) , TE=2.98ms, TR=5000ms, FA=4/5 degrees, TI=701/2500 ms. MRI collection used a 3T Siemens Prisma System with the body coil for signal transmission and a 64-channel head coil and 18-channel body coil for signal reception (software version VE11

Appendix A Table: summary of inclusion criteria for Alzheimer's cases (AD), Mild Cognitively Impaired (MCI) and Healthy controls (HC)

	HC	MCI	AD
ADNI1	- No Memory Complaints aside from those common to other normal subjects of that age range. - Normal memory function (Logical Memory II subscale (delayed Paragraph Recall) from the Wechsler Memory Scaled - Revised (the maximum score is 25): a) greater than or equal to 9 for 16 or more years of education b) greater than or equal to 5 for 8-15 years of education c) greater than or equal to 3 for 0-7 years of education. - MMSE score between 24 and 30 (inclusive) - CDR = 0. Memory Box score must be 0. - Cognitively normal, based on an absence of significant impairment in cognitive functions or activities of daily living.	- "Late MCI" - Memory complaint by subject or study partner that is verified by a study partner. - Abnormal memory function LM II subscale from WMSR a) less than or equal to 8 for 16 or more years of education b) less than or equal to 4 for 8-15 years of education c) less than or equal to 2 for 0-7 years of education. - MMSE between 24 and 30 (inclusive) - CDR = 0.5. Memory Box score must be at least 0.5. - General cognition and functional performance sufficiently preserved such that a diagnosis of Alzheimer's disease cannot be made by the site physician at the time of the screening visit.	"Mild AD" - Memory complaint by subject or study partner that is verified by a study partner. - Abnormal memory function LM II subscale from WMSR a) less than or equal to 8 for 16 or more years of education b) less than or equal to 4 for 8-15 years of education c) less than or equal to 2 for 0-7 years of education. - MMSE between 20 and 26 (inclusive) - Clinical Dementia Rating = 0.5, 1.0 - NINCDS/ADRDA criteria for probable AD.
ADNI2+GO		"EMCI" Memory complaint by subject or study partner that is verified by a study partner. - Abnormal memory function LM II subscale from WMSR a) 9-11 for 16 or more years of education b) 5-9 for 8-15 years of education c) 3-6 for 0-7 years of education. - MMSE between 24 and 30 (inclusive) - CDR = 0.5. Memory Box score must be at least 0.5. - General cognition and functional performance sufficiently preserved such that a diagnosis of Alzheimer's disease cannot be made by the site physician at the time of the screening visit	
ADNI 3	HC – same as ADNI1	EMCI – and MCI same as ADNI1	Mild AD – same as ADNI1

AIBL	Complaints aside from those common to other normal subjects of that age range. Not fulfilling criteria for MCI CDR of 0.5 possible for some individuas, 0 otherwise.	MCI - Winblad criteria - clinical review panel meetings. - personally, or through an informant, reported memory difficulties. - clinical diagnosis of MCI (i.e. previously diagnosed by a clinician) were further required to demonstrate a score 1.5 SD or more below the age-adjusted mean on at least one neuropsychological task applied at the time of the AIBL assessment in order to be retained in the MCI category. - Individuals who volunteered to take part as healthy controls had to fulfill more stringent criterion of impairment on two or more cognitive tests at a level at least 1.5 SD below the age-adjusted mean, in addition to having reported memory difficulties, to be classified as MCI. - NINCDS-ADRDA AD diagnosis (probable or possible) and MCI classifications were applied.	- clinical review panel meetings, leading to consensus diagnosis based on DSM4 and ICD10. - MMSE score <28, - failure on the Logical Memory test (as per ADNI criteria) - other evidence of possibly significant cognitive difficulty on neuropsychological testing - CDR score of 0.5 or greater - medical history suggestive of the presence of illnesses likely to impair cognitive function - an informant or personal history suggestive of impaired cognitive function. - NINCDS-ADRDA AD diagnosis (probable or possible) and MCI classifications were applied.
ARWIBO	Outpatients of the Neuroradiology Units, undergoing brain MR scan for reasons other than cognitive impairment. The prescription of MR for the following conditions was considered as an exclusion criterion: neurodegenerative diseases such as Alzheimer's, Parkinson's, progressive supranuclear palsy, Huntington's disease, multiple system atrophy, amyotrophic lateral sclerosis, cerebrovascular diseases such as stroke and TIA, and head trauma. MRI images also screened for structural findings (e.g. mass, aneurysm, white matter hyperintensities…).	MCI defined as presence of objective impairment in memory or other cognitive domains in the absence of functional impairment. Memory impairment defined as performance below the tenth percentile in at least one test among the Story Recall, Rey Auditory Verbal Learning Test, and Rey-Osterreith Complex Figure Recall (only when performance on the Copy is higher than the tenth percentile). Impairment in short-term memory is defined as performance below the tenth percentile in at least one among the Spatial Span and Digit Span. Impairment in language is defined as performance below the tenth percentile in at least one Verbal Fluency With Phonemic or Semantic Cues or Token Test. Impairment in nonverbal reasoning and frontal functions is defined as performance below the tenth percentile in at least one among Trail Making Test and Raven's Colored Progressive Matrices. Impairment in constructional praxis is defined as	AD was diagnosed according to NINCDS-ADRDA criteria

		performance below the tenth percentile on the Rey-Osterreith Complex Figure Copy.	
MAS	CN definition: Participants were classified as cognitively normal if performance on all test measures was above the 6.68 percentile (−1.5 SDs) or equivalent score compared to normative published values, they were not demented (see criterion (c) above) and they had normal function or minimal impairment in IADLs defined by a total average score <3.0 on the Bayer ADL scale. Cognitive complaint was allowed. Exclusion criteria - previous diagnosis of dementia, psychotic symptoms or a diagnosis of schizophrenia or bipolar disorder, multiple sclerosis, motor neuron disease, developmental disability, progressive malignancy (active cancer or receiving treatment for cancer, other than prostate – non-metastasized, and skin cancer) -or if they had medical or psychological conditions that may have prevented them from completing assessments. - MMSE score <24 adjusted for age, education and non-English speaking background at study entry - if they received a diagnosis of dementia after comprehensive assessment.	MCI definition - Winblad: *all* of the following criteria were met: (a) complaint of decline in memory or other cognitive function which may be self- or informant-reported (b) cognitive impairment on objective testing, i.e. not normal for age as determined by performance on at least one test measure 1.5 SDs or more *below* published normative values (or comparable standardized score provided in the normative source compared to age and/or education-matched samples) (c) not demented – participants did not have a pre-existing diagnosis of dementia on entry to the study, had an adjusted MMSE score of ≥24 and did not meet DSM-IV criteria for possible or probable dementia based on comprehensive clinical, cognitive and informant data gathered at assessment; (d) essentially normal function or minimal impairment in instrumental activities of daily living (IADLs) defined by a total average score <3.0 on the Bayer ADL Scale. Exclusion criteria as per controls.	NA
OATS		MCI defined using standard criteria, using both subjective reports and objective measures (Petersen, 2004; Winblad et al., 2004). MCI participants demonstrated cognitive impairment in one or more cognitive domains during the cognitive assessment. The diagnoses of MCI were made at consensus meetings attended by four or more investigators.	By consensus from clinical team, participants were diagnosed with dementia according to the DSM-IV criteria. Access to Medicare records obtained from the Health Insurance Commission of Australia after obtaining informed consent.
OASIS3	CN: Individuals generally healthy, cognitively normal (CDR = 0), with or without a family history of dementia.	MCI: CDR 0.5 very mild impairment, CDR 1 mild impairment. During the assessment, clinicians completed a diagnostic impression intake and Interview.	Dementia status assessed using clinical assessment protocols in accordance UDS (Alzheimer Coordinating Center Uniform Data Set). Dementia status used CDR Scale, CDR 2 indicating moderate

			dementia. During the assessment, clinicians completed a diagnostic impression intake and Interview.
MEMENTO	NA	MCI defined as (1) performing 1 SD worse than the subject's own age, sex and education-level group mean in one or more cognitive domains, this deviation being identified for the first time through cognitive tests performed recently (less than 6 months preceding screening phase), and (2) CDR ≤0.5 and not being demented	Caseness of AD based on DSM-IV criteria for dementia and NINCDS-ADRDA), and were reviewed by an independent committee.
PISA	HC – no significant neurological disorder (AD, stroke, vascular dementia, Parkinson's disease, Huntington's disease, normal-pressure hydrocephalus, CNS tumour or infection, seizure disorder, multiple sclerosis, or history of significant head trauma followed by haematoma, persistent neurological deficits or known structural brain abnormalities). N = 177 High genetic risk (*APOE* ε4 + ve AND/OR top quintile of PRS no-*APOE*), N = 56 Low genetic risk (*APOE* ε4 -ve AND lowest quintile of PRS no-*APOE*). European ancestry.	MCI: Meet DSM-5 Criteria for Mild Neurocognitive Disorder or NIA-AA Criteria for Mild Cognitive Impairment. Mini-mental state examination (MMSE) > 20 and a Clinical Dementia Rating (CDR) of 0.5 or 1.0. Clinical consensus meeting is held to establish the blinded clinical diagnosis based on current criteria for MCI and AD	AD: Meet DSM-5 Criteria for or Major Neurocognitive Disorder of the Alzheimer's Type or NIA-AA Criteria for Dementia of the Alzheimer's type. Mini-mental state examination (MMSE) > 20 and a Clinical Dementia Rating (CDR) of 0.5 or 1.0. Clinical consensus meeting is held to establish the blinded clinical diagnosis based on current criteria for MCI and AD

Appendix B. Conversion of effect sizes – from mean differences to cohen's d.

The cohen's d we report in the manuscript (Figure 3) correspond to the association effect size b_1 from the following model. We omit the covariates for simplicity, but they should not change the derivations.

$$\boldsymbol{ADvsHC} = b_1 \,.\, \boldsymbol{HippoVol_{STD}} + \boldsymbol{\varepsilon}$$

Where $ADvsHC$ is the case control status, and $\boldsymbol{HippoVol_{STD}}$ is the standardised hippocampal volume (mean 0, variance 1). In this case, each increase of 1 $\boldsymbol{HippoVol_{STD}}$ corresponds to an increase of 1 SD in hippocampal volume, hence b_1 is the cohen's d.

Most of the previous articles have reported mean differences in Hippocampal volume between cases and controls(Frisoni et al., 2008; Schuff et al., 2009; Vijayakumar & Vijayakumar, 2012). In practice, this is equivalent to fitting:

$$\boldsymbol{HippoVol} = b_2 \,.\, \boldsymbol{ADvsHC} + \boldsymbol{\varepsilon}$$

Here b_2 quantifies the mean difference in hippocampal volume between the AD and HC groups.

We can use the fact that :

$$b_1 = \frac{cov(\boldsymbol{ADvsHC}, \boldsymbol{HippoVol_{STD}})}{var(\boldsymbol{HippoVol_{STD}})} = \frac{cov(\boldsymbol{ADvsHC}, \boldsymbol{HippoVol})}{sd(\boldsymbol{HippoVol})}$$

And :

$$b_2 = \frac{cov(\boldsymbol{ADvsHC}, \boldsymbol{HippoVol})}{var(\boldsymbol{ADvsHC})}$$

To convert the mean group differences (b_2) into cohen's d (b_1) using the relationship:

$$b_1 = \frac{b_2 . var(\boldsymbol{ADvsHC})}{sd(\boldsymbol{HippoVol})}$$

We have converted into cohen's d, the mean differences reported in several studies, using the $sd(\boldsymbol{HippoVol})$ reported in the publications and the var(ADvsHC) that correspond to our meta-analysis (var(ADvsHC)=0.16). We obtained cohen's d ranging from 0.26-0.36. In comparison, our estimates were -0.18 (SE=0.024) for left hippocampus volume and -0.17 (SE=0.017) for right hippocampus volume, which corresponds to differences of 15-20% between volumes of cases and controls (based on converting cohen's d to mean differences). Our smaller effect sizes may be attributable to winner's curse and small sample bias that tend to inflate the results previously published. In addition, previous reports did not systematically control for differences in age, sex and ICV between cases and controls, which can also lead to inflated differences.

	N	HC	AD	sd(HV)	b2 (mean HV difference)	Pct difference	b1 (Cohen's d)
Schuff et al, Hippocampal volume left-right average (mm3)	223	2133	1631	303.4	-502	-23%	-0.26
Frisoni et al, hippo volume (mm3) - Left	36	3889	2705	694	-1184	-30%	-0.27
Right	36	4216	2745	709.5	-1471	-34%	-0.33
Vijayakumar et al., hippo volume (mm3) - left	26	2540	1846	308.1	-694	-27%	-0.36
right	26	2660	1997	353.8	-663	-25%	-0.29

Appendix B table : Cohen's d of the associations between hippocampal volume and AD status, previously reported in the literature.

Dufouil, C., Dubois, B., Vellas, B., Pasquier, F., Blanc, F., Hugon, J., Hanon, O., Dartigues, J.-F., Harston, S., Gabelle, A., Ceccaldi, M., Beauchet, O., Krolak-Salmon, P., David, R., Rouaud, O., Godefroy, O., Belin, C., Rouch, I., Auguste, N., … MEMENTO cohort Study Group. (2017). Cognitive and imaging markers in non-demented subjects attending a memory clinic: Study design and baseline findings of the MEMENTO cohort. *Alzheimer's Research & Therapy*, *9*(1), 67. https://doi.org/10.1186/s13195-017-0288-0

Ellis, K. A., Bush, A. I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N. T., Lenzo, N., Martins, R. N., Maruff, P., Masters, C., Milner, A., Pike, K., Rowe, C., Savage, G., Szoeke, C., Taddei, K., Villemagne, V., Woodward, M., … AIBL Research Group. (2009). The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *International Psychogeriatrics*, *21*(4), 672–687. https://doi.org/10.1017/S1041610209009405

Frisoni, G. B., Ganzola, R., Canu, E., Rüb, U., Pizzini, F. B., Alessandrini, F., Zoccatelli, G., Beltramello, A., Caltagirone, C., & Thompson, P. M. (2008). Mapping local hippocampal changes in Alzheimer's disease and normal ageing with MRI at 3 Tesla. *Brain*, *131*(12), 3266–3276. https://doi.org/10.1093/brain/awn280

Frisoni, G. B., Molinuevo, J. L., Altomare, D., Carrera, E., Barkhof, F., Berkhof, J., Delrieu, J., Dubois, B., Kivipelto, M., Nordberg, A., Schott, J. M., van der Flier, W. M., Vellas, B., Jessen, F., Scheltens, P., & Ritchie, C. (2020). Precision prevention of Alzheimer's and other dementias: Anticipating future needs in the control of risk factors and implementation of disease-modifying therapies. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *16*(10), 1457–1468. https://doi.org/10.1002/alz.12132

Kochan, N. A., Slavin, M. J., Brodaty, H., Crawford, J. D., Trollor, J. N., Draper, B., & Sachdev, P. S. (2010). Effect of Different Impairment Criteria on Prevalence of "Objective" Mild Cognitive Impairment in a Community Sample. *The American Journal of Geriatric Psychiatry*, *18*(8), 711–722. https://doi.org/10.1097/JGP.0b013e3181d6b6a9

Koncz, R., Mohan, A., Dawes, L., Thalamuthu, A., Wright, M., Ames, D., Lee, T., Trollor, J., Wen, W., & Sachdev, P. (2018). Incidental findings on cerebral MRI in twins: The Older Australian Twins Study. *Brain Imaging and Behavior*, *12*(3), 860–869. https://doi.org/10.1007/s11682-017-9747-2

LaMontagne, P. J., Keefe, S., Lauren, W., Xiong, C., Grant, E. A., Moulder, K. L., Morris, J. C., Benzinger, T. L. S., & Marcus, D. S. (2018). OASIS-3: Longitudinal neuroimaging, clinical and cognitive dataset for normal aging and Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *14*(7), P1097. https://doi.org/10.1016/j.jalz.2018.06.1439

Lindroth, H., Nair, V. A., Stanfield, C., Casey, C., Mohanty, R., Wayer, D., Rowley, P., Brown, R., Prabhakaran, V., & Sanders, R. D. (2019). Examining the identification of age-related atrophy between T1 and T1 + T2-FLAIR cortical thickness measurements. *Scientific Reports*, *9*(1), 11288. https://doi.org/10.1038/s41598-019-47294-2

Lorenzini, L., Ingala, S., Wink, A. M., Kuijer, J. P., Wottschel, V., Dijsselhof, M., Sudre, C. H., Haller, S., Molinuevo, J. L., Gispert, J. D., Cash, D. M., Thomas, D. L., Vos, S. B., Prados, F., Petr, J., Wolz, R., Palombit, A., Schwarz, A. J., Gael, C., … for the EPAD consortium. (2021). *The European Prevention of Alzheimer's Dementia (EPAD) MRI Dataset and Processing Workflow* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.09.29.462349

Lupton, M. K., Robinson, G. A., Adam, R. J., Rose, S., Byrne, G. J., Salvado, O., Pachana, N. A., Almeida, O. P., McAloney, K., Gordon, S. D., Raniga, P., Fazlollahi, A., Xia, Y., Ceslis, A., Sonkusare, S., Zhang, Q., Kholghi, M., Karunanithi, M., Mosley, P. E., … Breakspear, M. (2020). A prospective cohort study of prodromal Alzheimer's disease: Prospective Imaging Study of Ageing: Genes, Brain and Behaviour (PISA). *NeuroImage : Clinical*, *29*, 102527. https://doi.org/10.1016/j.nicl.2020.102527

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., Bartsch, A. J., Jbabdi, S., Sotiropoulos, S. N., Andersson, J. L. R., Griffanti, L., Douaud, G., Okell, T. W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., … Smith, S. M. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, *19*(11), Article 11. https://doi.org/10.1038/nn.4393

Régy, M., Dugravot, A., Sabia, S., Fayosse, A., Mangin, J.-F., Chupin, M., Fischer, C., Bouteloup, V., Dufouil, C., Chêne, G., Paquet, C., Hanseeuw, B., Singh-Manoux, A., & Dumurgier, J. (2022). Association of APOE ε4 with cerebral gray matter volumes in non-demented older adults: The MEMENTO cohort study. *NeuroImage*, *250*, 118966. https://doi.org/10.1016/j.neuroimage.2022.118966

Riello, R., Sabattoli, F., Beltramello, A., Bonetti, M., Bono, G., Falini, A., Magnani, G., Minonzio, G., Piovan, E., Alaimo, G., Ettori, M., Galluzzi, S., Locatelli, E., Noiszewska, M., Testa, C., & Frisoni, G. B. (2005). Brain volumes in healthy adults aged 40 years and over: A voxel-based morphometry study. *Aging Clinical and Experimental Research*, *17*(4), 329–336. https://doi.org/10.1007/BF03324618

Sachdev, P. S., Brodaty, H., Reppermund, S., Kochan, N. A., Trollor, J. N., Draper, B., Slavin, M. J., Crawford, J., Kang, K., Broe, G. A., Mather, K. A., Lux, O., & Memory and Ageing Study Team. (2010). The Sydney Memory and Ageing Study (MAS):

Methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70-90 years. *International Psychogeriatrics*, *22*(8), 1248–1264. https://doi.org/10.1017/S1041610210001067

Sachdev, P. S., Lammel, A., Trollor, J. N., Lee, T., Wright, M. J., Ames, D., Wen, W., Martin, N. G., Brodaty, H., Schofield, P. R., & OATS research team. (2009). A comprehensive neuropsychiatric study of elderly twins: The Older Australian Twins Study. *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, *12*(6), 573–582. https://doi.org/10.1375/twin.12.6.573

Sachdev, P. S., Lee, T., Wen, W., Ames, D., Batouli, A. H., Bowden, J., Brodaty, H., Chong, E., Crawford, J., Kang, K., Mather, K., Lammel, A., Slavin, M. J., Thalamuthu, A., Trollor, J., Wright, M. J., & OATS Research Team. (2013). The contribution of twins to the study of cognitive ageing and dementia: The Older Australian Twins Study. *International Review of Psychiatry (Abingdon, England)*, *25*(6), 738–747. https://doi.org/10.3109/09540261.2013.870137

Schuff, N., Woerner, N., Boreta, L., Kornfield, T., Shaw, L. M., Trojanowski, J. Q., Thompson, P. M., Jack, C. R., Jr, Weiner, M. W., & the Alzheimer's; Disease Neuroimaging Initiative. (2009). MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. *Brain*, *132*(4), 1067–1077. https://doi.org/10.1093/brain/awp007

Solomon, A., Kivipelto, M., Molinuevo, J. L., Tom, B., Ritchie, C. W., & EPAD Consortium. (2019). European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD LCS): Study protocol. *BMJ Open*, *8*(12), e021017. https://doi.org/10.1136/bmjopen-2017-021017

Tsang, R. S. M., Sachdev, P. S., Reppermund, S., Kochan, N. A., Kang, K., Crawford, J., Wen, W., Draper, B., Trollor, J. N., Slavin, M. J., Mather, K. A., Assareh, A., Seeher,

K. M., & Brodaty, H. (2013). Sydney Memory and Ageing Study: An epidemiological cohort study of brain ageing and dementia. *International Review of Psychiatry (Abingdon, England)*, *25*(6), 711–725. https://doi.org/10.3109/09540261.2013.860890

Vijayakumar, A., & Vijayakumar, A. (2012). Comparison of Hippocampal Volume in Dementia Subtypes. *ISRN Radiology*, *2013*, 174524. https://doi.org/10.5402/2013/174524