Supplementary Material

Relationship Between Cerebrospinal Fluid Alzheimer's Disease Biomarker Values Measured via Lumipulse Assays and Conventional ELISA: Single-Center Experience and Systematic Review

Overview of QUABICS

QUABICS (QUality Assessment tool for BIomarker measurement procedure Comparison Studies) was designed specifically to assess the quality of biomarker measurement procedure comparison studies for systematic review. It is a modification of BIOCROSS [1], a quality assessment tool for biomarker-based cross-sectional studies.

QUABICS has 10 items. Each is scored from 0 to 2, thus the scoring range is 0 to 20.

Scoring system of QUABICS

1. Study Rational:
   2 points: Clearly describing the rationale of the study assessing biomarker measurement procedure comparison.
   0 point: None described.
   1 point: Some rational written (not 0 or 2 points).

2-1. Study population selection
   2 points: Consecutive recruitment based on criteria.
   0 point: No description of the method of population selection.
   1 point: Method described but not consecutive recruitment (ex. case-control etc.) (not 0 or 2 points)

2-2. Study population representativeness
   2 points: Recruitment from a real-world setting with a recruitment rate ≥ 50%.
   0 point: Less than 50 participants.
   1 point: More than 50 participants but not fulfilling 2 points (not 0 or 2 points) (ex. assessed only in a population biased from real-world setting).

2-3. Study population characteristics
   2 points: Characteristics of study population well described (age, sex, recruited country, diagnosis, etc.)
   0 point: None described.
   1 point: Some characteristics described but insufficient (not 0 or 2 points).

3-1. Sampling and handling of specimens
   2 points: Adequate according to a standard guideline (ex. "The Alzheimer's Association international guidelines for handling of cerebrospinal fluid for routine clinical

measurements of amyloid β and tau" published in 2021 [2] for CSF AD biomarker studies)

0 point: Inadequate according to a standard guideline or no method described.

1 point: Not 0 or 2 points.

3-2. Detailed description of the measurement protocol

2 points: Detailed description of protocol adequate for replication for both measurements.

0 point: Insufficient description of the protocol for both measurements.

1 point: Not 0 or 2 points (ex. detailed description for one measurement but not the other).

3-3. Quality control of each measurement

2 points: Adequate quality control results documented in the manuscript for both measurements (ex. coefficient of variation (CV) ≤ 10%)

0 point: No quality control results reported.

1 point: Not 0 or 2 points (ex. adequate quality control results for one measurement but not the other).

4-1. Statistical method

2 points: Nonparametric analysis assuming variability in both axes (ex. Passing–Bablok regression)

0 point: Parametric AND assuming variability only in the y-axis (ex. only using Pearson's correlation) or no method description

1 point: Not 0 or 2 points (ex. Deming regression).

4-2. Data reporting

2 points: Reporting median (mean) AND confidence interval of slope and intercept.

1 point: Reporting median (mean) slope/intercept AND p values.

0 point: not 1 or 2 points (ex. reporting only median (mean), reporting only correlation coefficient).

5. Data interpretation

2 points: Discussion for correlation AND proportional/systematic biases referring to previous reports and limitations of the study.

0 point: No discussion for biomarker-correlation.

1 point: Not 0 or 2 points (ex. referring to the degree of correlation, but not to biases of the actual measurement results).

Inter-rater agreement

The scores for 13 studies were evaluated independently by two authors (M.K and S.K). The total scores showed a moderate to good correlation between the two raters (Supplementary Figure 3A). The frequency of score matched between two raters for each item is provided in Supplementary Figure 3B and was 62–100%.

Interpretation of QUABICS total scores in this study

There is no unified cut-off score for including or excluding studies for systematic reviews, and the cut-off should be determined cautiously in each review. In this study, based on histogram (Supplementary Figure 4), we used a provisional category listed below.

High quality: 16–20

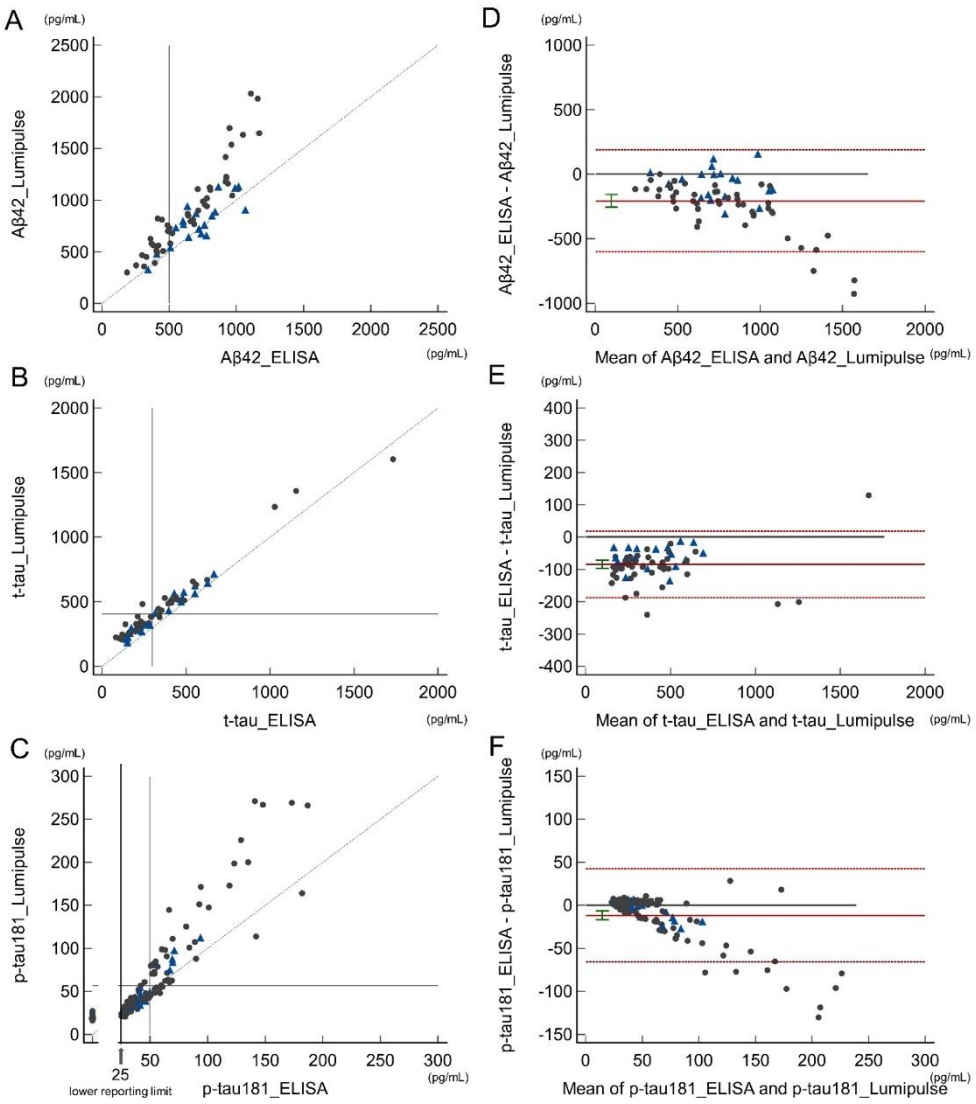
Medium quality: 12–15

Low quality: 0–11

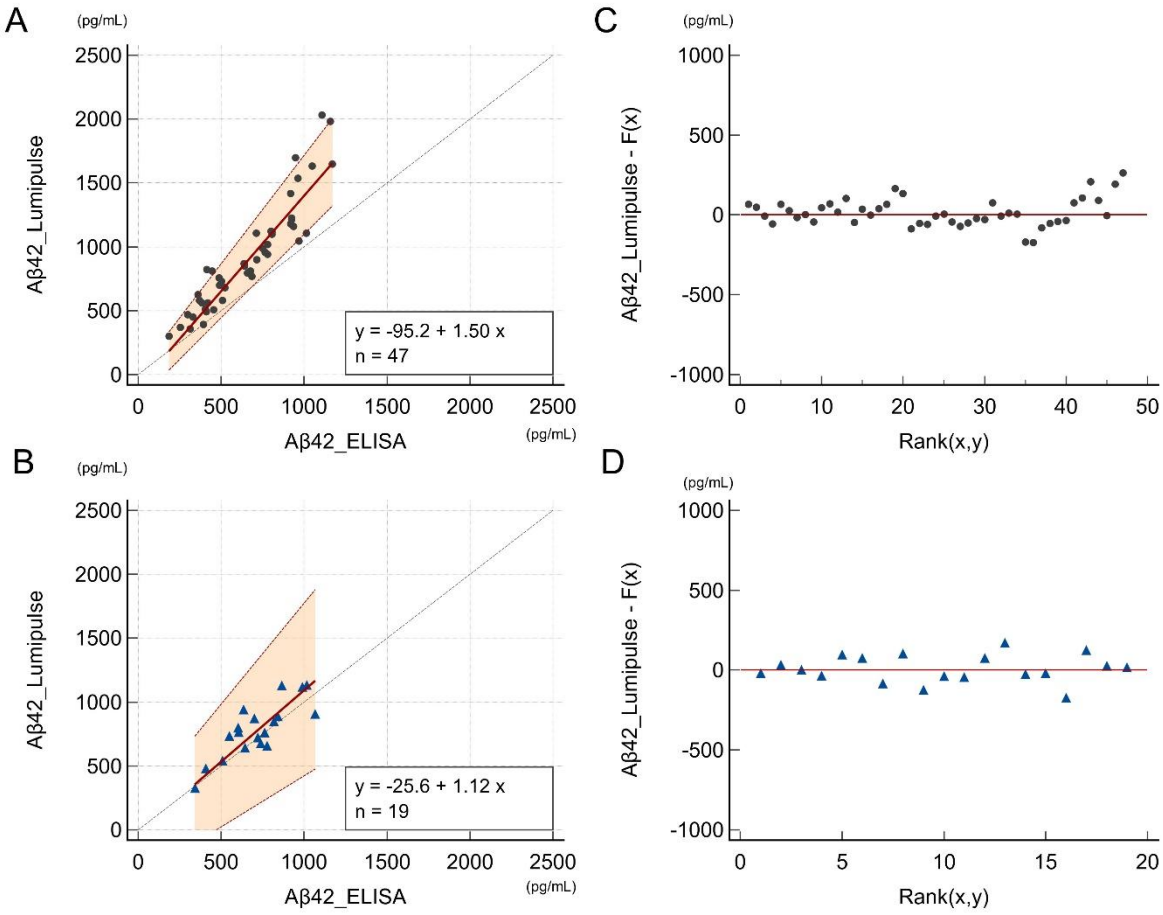
REFERENCES

[1] Wirsching J, Graßmann S, Eichelmann F, Harms LM, Schenk M, Barth E, Berndzen A, Olalekan M, Sarmini L, Zuberer H, Aleksandrova K (2018) Development and reliability assessment of a new quality appraisal tool for cross-sectional studies using biomarker data (BIOCROSS). *BMC Med Res Methodol* **18**, 122.

[2] Hansson O, Batrla R, Brix B, Carrillo MC, Corradini V, Edelmayer RM, Esquivel RN, Hall C, Lawson J, Bastard NL, Molinuevo JL, Nisenbaum LK, Rutz S, Salamone SJ, Teunissen CE, Traynham C, Umek RM, Vanderstichele H, Vandijck M, Wahl S, Weber CJ, Zetterberg H, Blennow K (2021) The Alzheimer's Association international guidelines for handling of cerebrospinal fluid for routine clinical measurements of amyloid beta and tau. *Alzheimers Dement* **17**, 1575-1582.



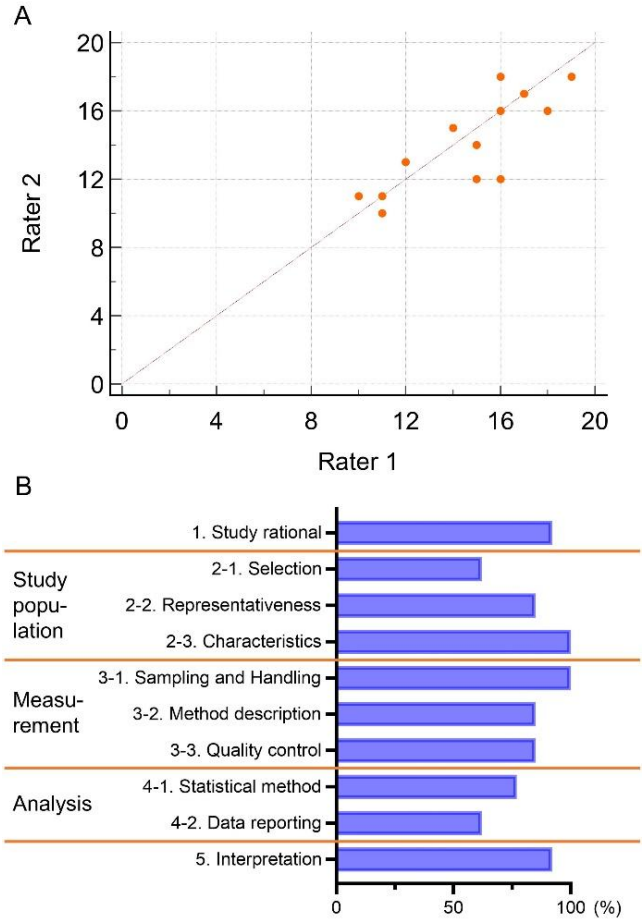
Supplementary Figure 1. Scatter and Bland–Altman plots for $A\beta_{42}$, t-tau, and p-tau181 using Lumipulse assays and ELISA. Each point represents the measurement result for each participant. Scatter plots for $A\beta_{42}$ (A, n = 66), t-tau (B, n = 66), and p-tau181 (C, n = 135) are shown. Lumipulse measurements were conducted in CSF samples directly collected in recommended polypropylene low-binding tubes in results represented in gray circles and were conducted in those aliquoted as biobank samples (same as those used in ELISA for $A\beta_{42}$ and t-tau) in results represented in blue triangles (n = 19). Scatter plots for $A\beta_{42}$ (A), t-tau (B), and p-tau181 (C) are shown. Diagonal dotted lines represent lines or equality (x = y). Horizontal and vertical lines represent predetermined cut-offs. p-tau181 ELISA values were below the lower reporting limit in 17 patients. Bland–Altman difference plots for $A\beta_{42}$ (D), t-tau (E), and p-tau181 (F) are shown. $A\beta_{42}$ (D) and p-tau181 (F) were lower in ELISA in the higher range suggesting a proportional difference. While no difference between different sampling procedures was suspected for t-tau and p-tau181, a bias was observed towards lower $A\beta_{42}$ measured by Lumipulse using aliquoted biobank samples compared to measurement results using directly collected samples. $A\beta_{42}$, amyloid-β 1-42; t-tau, total tau; p-tau 181, tau phosphorylated at threonine 181; ELISA, enzyme-linked immunosorbent assay.



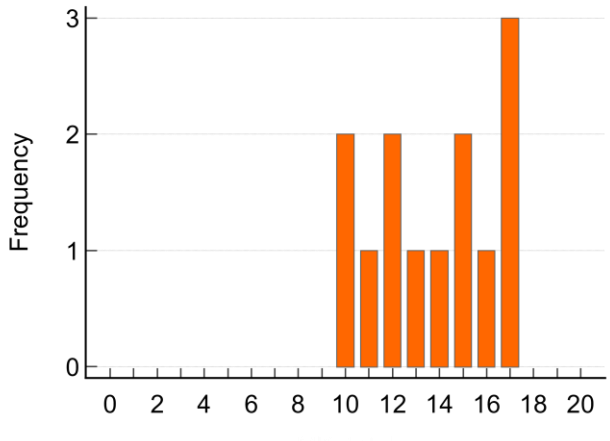
E

	Slope	Intercept	± 1.96 RSD Interval
All	1.41 [1.23 to 1.60]	-77.8 [-198.4 to 44.5]	-211.5 to 211.5
Circle	1.50 [1.30 to 1.70]	-95.2 [-207.0 to 19.3]	-176.9 to 176.9
Triangle	1.12 [0.80 to 1.58]	-21.1 [-376.4 to 193.6]	-175.8 to 175.8

Supplementary Figure 2. Passing–Bablok analyses for $A\beta_{42}$ differentiating sampling procedures. Each point represents the measurement result of each participant. Lumipulse measurements were conducted in CSF samples directly collected in recommended polypropylene low-binding tubes in results represented in gray circles (A, C), and were conducted in those aliquoted as biobank samples (same as ELISA) in results represented in blue triangles (B, D, n = 19). Passing–Bablok regression lines with 95% confidence intervals (A, B) and residuals from the regression line plotted by rank order (C, D) are shown for $A\beta_{42}$. (E) Median [95%CI] of slopes and intercepts, and ± 1.96 residual standard deviation (RSD) intervals are summarized. Measurement results using different sampling procedures (circle) showed a slope higher than 1.0. Although underpowered, measurement results using the same sampling procedure (triangle) tended to show a lower slope close to 1.0 (B, E).

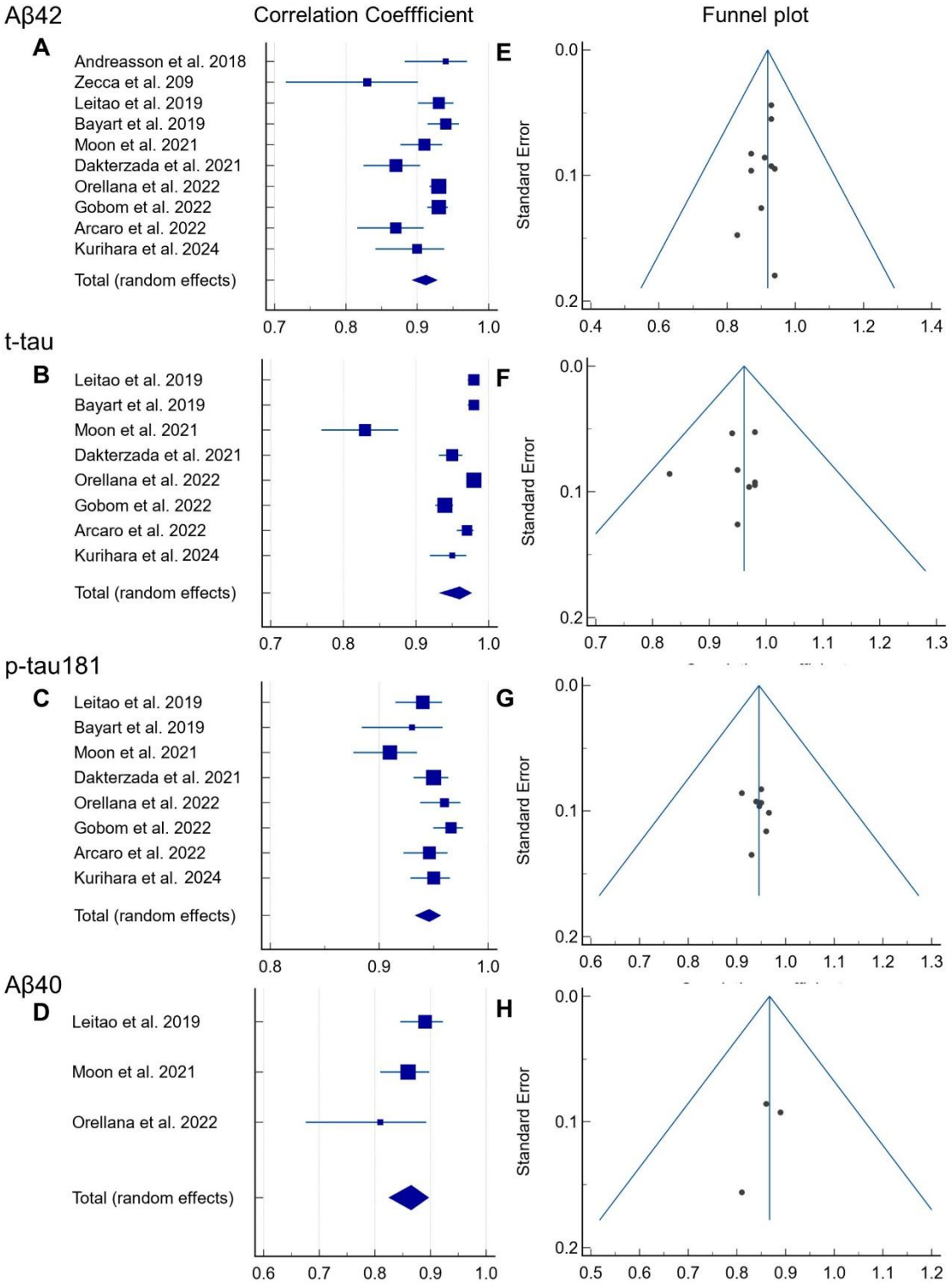


Supplementary Figure 3. Inter-rater agreement of QUABICS scores. A) The scatter plot of total scores rated by two authors. The total scores showed a moderate to good correlation between the two raters (Spearman's correlation coefficient 0.83 [0.51–0.95]). B) The frequency of score matched between two raters for each item. The frequency of score matched was 62–100%.



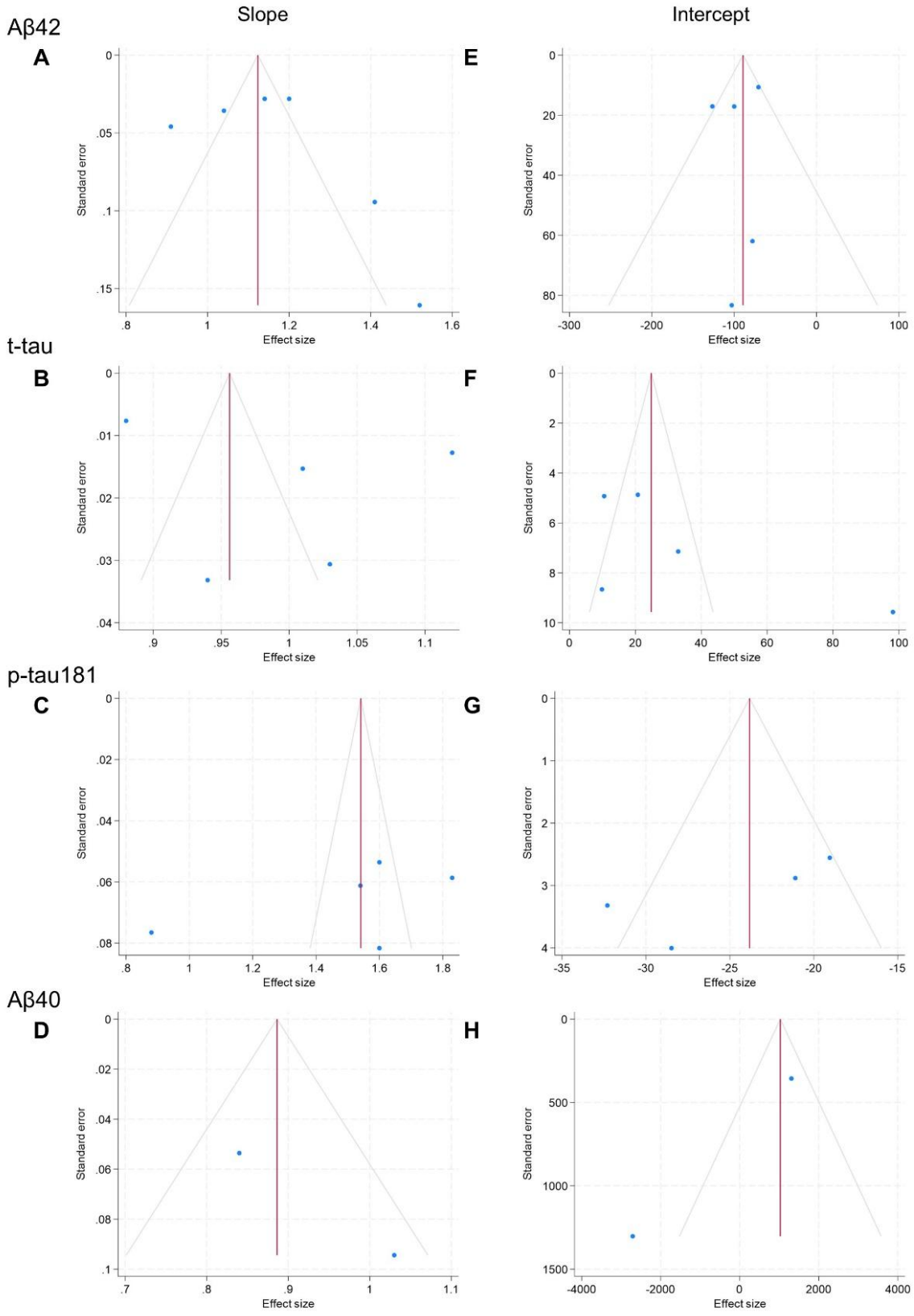
Supplementary Figure 4. Histogram of QUABICS total scores of the studies in this

systematic review. The total scores of the 13 studies after discussion by two raters are summarized.



Supplementary Figure 5. Forrest plot, meta-analysis results, and Funnel plot of studies reporting correlation coefficients for each biomarker. Studies reporting correlation coefficients based on log-transformed values were excluded. Pooled correlation coefficients were 0.91

(95%CI 0.89 to 0.93) for $A\beta_{42}$ (A), 0.96 (95%CI 0.93 to 0.98) for t-tau (B), 0.95 (95%CI 0.93 to 0.96) for p-tau181 (C), and 0.87 (95%CI 0.83 to 0.90) for $A\beta_{40}$ (D). Funnel plot showed no asymmetry suggesting minimal publication bias.



Supplementary Figure 6. Funnel plot of studies reporting slope and intercepts for each biomarker. Studies with larger standard error tended to report higher slopes in $A\beta_{42}$ (A). Otherwise Funnel plots showed no obvious asymmetry suggesting reporting bias for $A\beta_{42}$, t-tau, p-tau181 (B, C, E, F, G). The sample size was too small to assess publication bias for $A\beta_{40}$ (D, H).