Supplementary Materials

Progression subtypes in Parkinson’s disease identified by a data driven multi cohort analysis

Tom Hähnel[1,2*], Tamara Raschka[1,3] Stefano Sapienza[4,5], Jochen Klucken[4,5,6], Enrico Glaab[4], Jean-Christophe Corvol[7], Björn H. Falkenburger[2,8], Holger Fröhlich[1,3]

1. Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin, Germany
2. Department of Neurology, Medical Faculty and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Dresden, Germany
3. Bonn-Aachen International Center for IT, University of Bonn, Bonn, Germany
4. Biomedical Data Science, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg
5. Luxembourg Institute of Health (LIH), Strassen, Luxembourg
6. Centre Hospitalier de Luxembourg (CHL), Luxembourg
7. Sorbonne Université, Paris Brain Institute – ICM, Inserm, CNRS, Assistance Publique Hôpitaux de Paris, Pitié-Salpêtrière Hospital, Department of Neurology, Paris, France
8. German Center for Neurodegenerative Diseases (DZNE), Tatzberg 41, 01307 Dresden, Germany

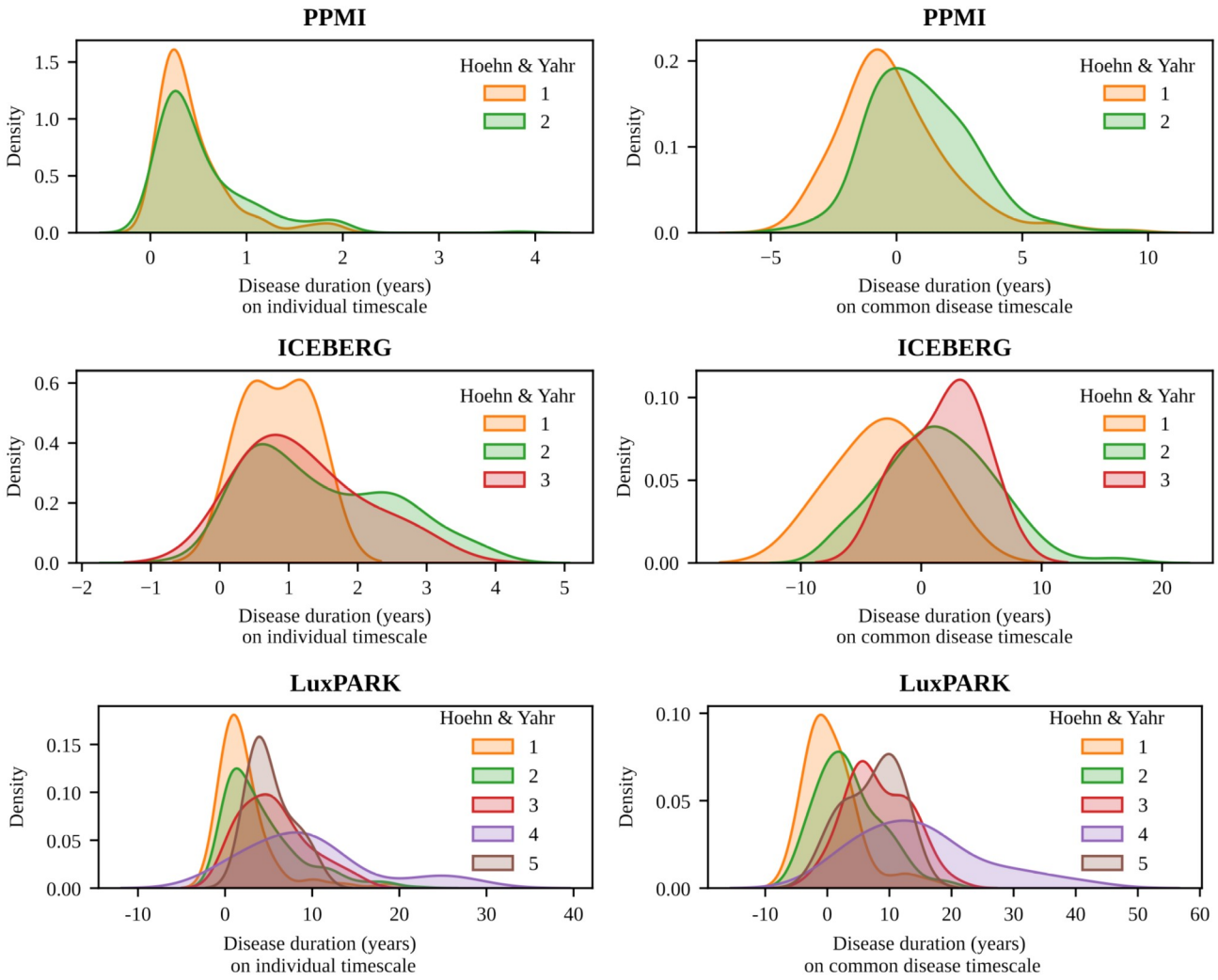
Corresponding author:
Dr. med. Tom Hähnel
Department of Bioinformatics
Fraunhofer Institute for Algorithms and Scientific Computing (SCAI)
Schloss Birlinghoven 1
53757 Sankt Augustin, Germany
Phone: +49 (0)351 458 11880
Email: tom.haehnel@scai-extern.fraunhofer.de

Running title: Parkinson's Disease Progression Subtypes

Progression subtypes in Parkinson’s disease identified by a data driven multi cohort analysis

Table of Contents

Supplementary Figures and Tables



Supplementary Figure 1: Effect of time-aligning PwPD on distributions of H&Y stages
The H&Y distributions from PPMI, ICEBERG and LuxPARK at baseline are depicted. On the left side, H&Y stages are plotted against the original time scale. On the right side, H&Y stages are plotted against the common timescale calculated from the LTJMM.
Abbreviations: H&Y: Hoehn&Yahr, LTJMM: latent time joint mixed-effects model

Study and parameter	Fast-progressing subtype	Slow-progressing subtype	p-value
Age (years)			
PPMI	67.6	62.0	*<0.0001*
ICEBERG	64.0	63.7	0.86
LuxPARK	70.1	66.8	*<0.0001*
Age at onset (years)			
PPMI	67.1	61.4	*<0.0001*
ICEBERG	62.5	61.4	0.86
LuxPARK	67.0	60.5	*<0.0001*
Disease duration (years)			
PPMI	0.3	0.3	0.86
ICEBERG	1.3	1.2	0.86
LuxPARK	2.8	3.0	0.44
Sex (% male)			
PPMI	73.0 %	65.4 %	0.38
ICEBERG	71.4 %	58.9 %	0.38
LuxPARK	73.2 %	65.4 %	0.21

Supplementary Table 1: Demographic differences between progression subtypes

PwPD baseline characteristics for fast-progressing and slow-progressing subtypes for the three cohorts PPMI, ICEBERG and LuxPARK. For sex, percentage of male PwPD is shown. For other characteristics, median values are reported. Corresponding p-values were corrected for multiple testing using Benjamini-Hochberg procedure.

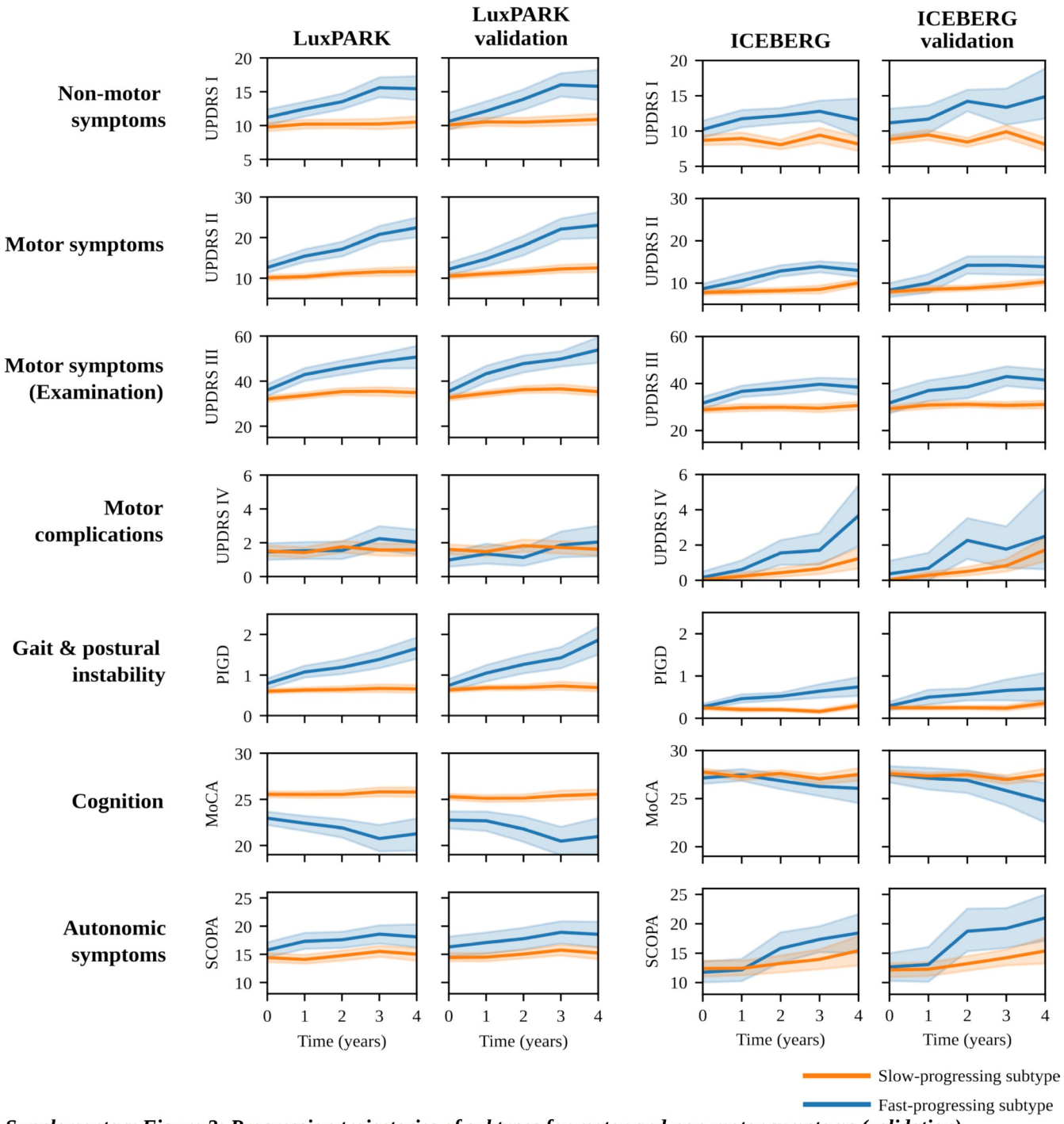
Symptom domain	Outcome	Definition/calculation of the outcome
Anxiety	NMSQ Anxiety	NMSQ item 17
	HADS anxiety	HADS anxiety sub-score
	STA	STA sum score
	PDQ39 Anxiety	PDQ39 item 21
	UPDRS I Anxiety	UPDRS I item 4
Apathy	DAS	DAS sum score
	SAS	SAS sum score
	UPDRS I Apathy	UPDRS I item 5
Autonomic symptoms	NMSQ Autonomic	NMSQ sum of items 4, 5, 6, 7, 8, 9, 19, 20, 28
	SCOPA-AUT	SCOPA sum score
	UPDRS I Autonomic	UPDRS I sum of items 10, 11, 12
Attention	NMSQ Attention	NMSQ item 15
	Letter Number Sequencing	Letter Number Sequencing score
	Digit Span	Digit Span Score: sum of forward scores and backward scores
	MATTIS Attention	MATTIS attention sub-score
	MOCA Attention	MoCA sum of: digits, letters, subtraction points
	MMSE Attention	MMSE attention sub-score
	PDQ39 Attention	PDQ39 item 31
	SIQCDE Attention	Short IQCODE score item 11
Conceptualization	MATTIS Conceptualization	MATTIS conceptualization sub-score
	MOCA Abstraction	MoCA abstraction sub-score
	FAB Conceptualization	FAB item 1
Language	Boston Naming Test	Boston Naming Test sum score
	MOCA Language + Naming	MoCA sum of items: naming, repeat and verbal fluency task
	VFT phonematic F	phonematic VFT F total word count
	VFT phonematic S	phonematic VFT S total word count
	VFT semantic animal	semantic VFT animal total word count
	VFT semantic sum	semantic VFT total word count (sum of tasks colors, fruits, towns, animals)
	VFT semantic supermarket	semantic VFT supermarket total word count
	MMSE Language	MMSE language sub-score
	FAB VFT	FAB lexical fluency item
Memory	NMSQ Memory	NMSQ item 12
	CERAD Words DR	CERAD word count immediate recall
	CERAD Words IR	CERAD word count delayed recall
	CERAD Words Recognition	CERAD recognition (number of correct, Yes + No)
	Hopkins Verbal Learning Test DR	Hopkins Verbal Learning Test delayed recall
	Hopkins Verbal Learning Test IR	Hopkins Verbal Learning Test immediate recall
	MOCA Orientation + Memory	MoCA sum of: memory (uncued only), orientation
	MATTIS Memory	MATTIS memory sub-score
	MMSE Memory	MMSE sum of: orientation (location + time), words memorization
	SIQCDE Memory	Short IQCODE score sum of items 1, 2, 3, 4, 5, 6, 7
	PDQ39 Memory	PDQ39 item 32
Overall Cognition	MATTIS	MATTIS sum score
	MMSE	MMSE sum score
	MoCA	MoCA sum score
	SIQCDE	Short IQCODE score sum score
	FAB	FAB sum score

Symptom domain	Outcome	Definition/calculation of the outcome
	PDQ39 Cognition	PDQ39 sum of items 31, 32
	UPDRS I Cognition	UPDRS I item 1
	NMSQ Cognition	NMSQ sum of items 12, 15
Visu-executive function	MMSE Construction	MMSE item 30
	Judgment Line Orientation	Judgment of Line Orientation sum score
	Symbol Digit Modalities	Symbol Digital Modalities sum score
	Stroop Errors	Stroop test number of errors
	Stroop Time	Stroop test required time
	Trailmaking A	Trail Making Test A time
	Trailmaking B	Trail Making Test B time
	MATTIS Initiation + Construction	MATTIS sum of: sub-score initiation, sub-score construction
	FAB 3-6	FAB sum of items 3, 4, 5, 6
	MOCA Visuspatial/Executive	MoCA visuospatial/executive sub-score
Depression	BDI	BDI sum score
	GDS	GDS sum score
	HADS depression	HADS depression sub-score
	PDQ39 Depression	PDQ39 sum of items 17, 18, 19, 20, 22
	NMSQ Depression	NMSQ sum of items 13, 16
	UPDRS I Depression	UPDRS 1 item 3
Fatigue	UPDRS I Fatigue	UPDRS 1 item 13
Hallucinations	NMSQ Hallucination	NMSQ sum of items 14, 30
	UPDRS I Hallucinations	UPDRS 1 item 2
Impulsivity	QUIP	QUIP sum score
	QUIP-RS	QUIPRS sum score
Motor symptoms (overall)	PDQ39 ADL	PDQ39 ADL sub-score
	Pegboard	PEGBoard sum of: average of left hand, right hand and both hands
	UPDRS II	UPDRS II sum score
	UPDRS III off	UPDRS III sum score (OFF only)
	UPDRS III on	UPDRS III sum score (ON only)
	UPDRS IV	UPDRS IV sum score
Non motor symptoms (overall)	NMSQ	NMSQ sum score
	UPDRS I	UPDRS I sum score
Overall disease severity	UPDRS I-III on	UPDRS I, II, III sum (ON only)
	UPDRS I-III off	UPDRS I, II, III sum (OFF only)
	FAQ	FAQ sum score
	PDQ39	PDQ39 sum score
	SEADL	SEADL score
	H&Y	Hoehn & Yahr
	CGIS	CGI-S score
Pain	NMSQ Pain	NMSQ item 10
	PDQ39 Pain	PDQ39 sum of items 37, 38
	UPDRS I Pain	UPDRS 1 item 9
Axial & PIGD symptoms	UPDRS III axial off	UPDRS III axial score (OFF only)
	UPDRS III axial on	UPDRS III axial score (ON only)
	FOGAC	FOGAC sum score
	FOGQ	FOGQ sum score

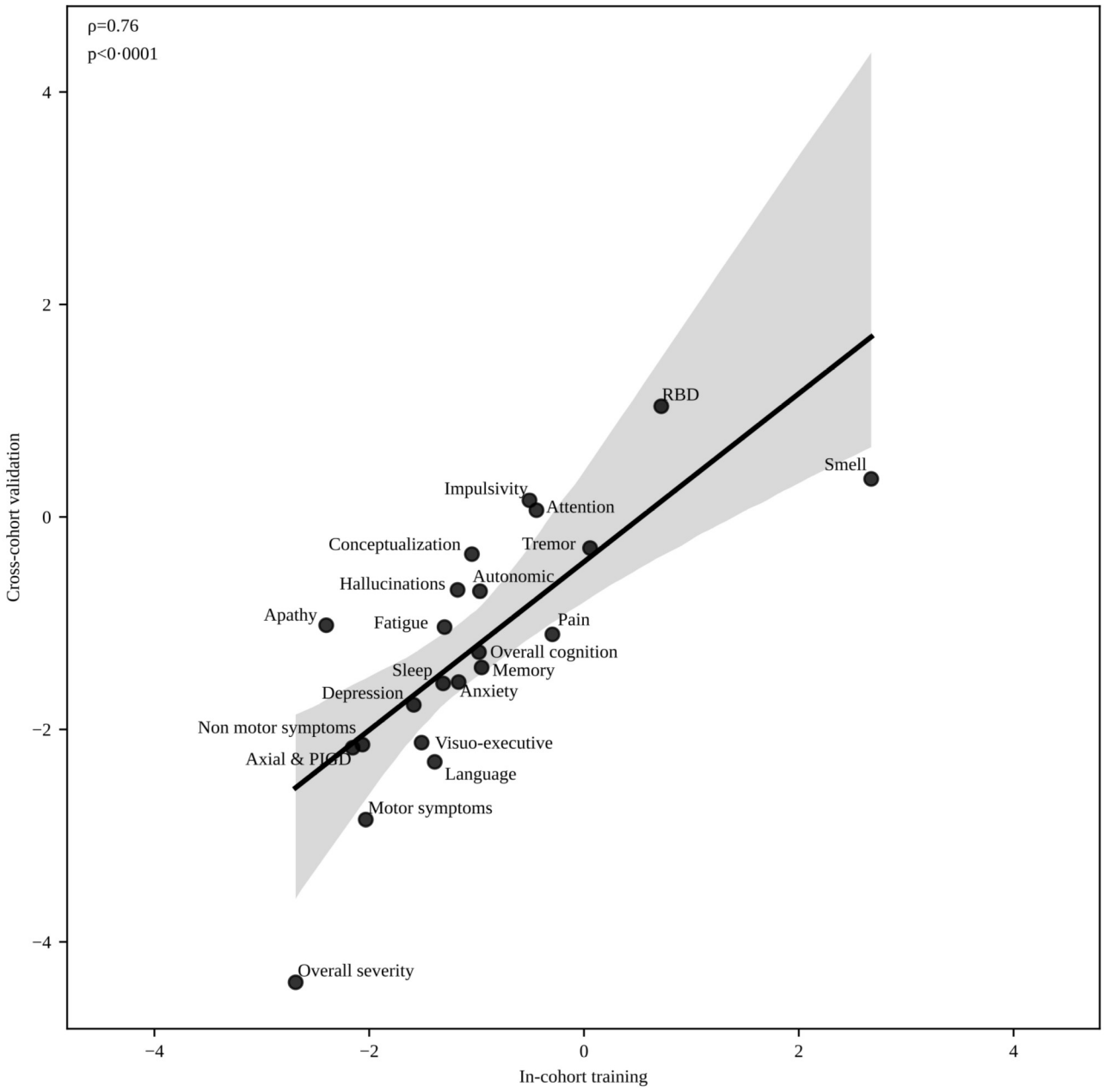
Symptom domain	Outcome	Definition/calculation of the outcome
	GABS Examination	GABS sum of items 8-24
	GABS Questionnaire	GABS sum of items 1-7
	NFOGQ	NFOGQ sum score
	PDQ39 Mobility	PDQ39 mobility sub-score
	PIGD off	PIGD score (OFF only)
	PIGD on	PIGD score (ON only)
	TUG	Timed Up and Go time
Sleep (general)	ESS	ESS sum score
	PDSS	PDSS sum score
	UPDRS I Sleep	UPDRS I sum of items 7, 8
	NMSQ Sleep	NMSQ sub of items 22, 23
RBD Sleep	RBD-HK	RBD-HK sum score
	RBD-SQ	RBD-SQ sum score
	NMSQ RBD	NMSQ sum of items 24, 25
Smell	NMSQ Smell	NMSQ item 2
	Sniffin Test	Sniffin Test score
	UPSIT	UPSIT sum score
Tremor	TD off	TD score (OFF only)
	TD on	TD score (ON only)

Supplementary Table 2: Construction of symptom domains

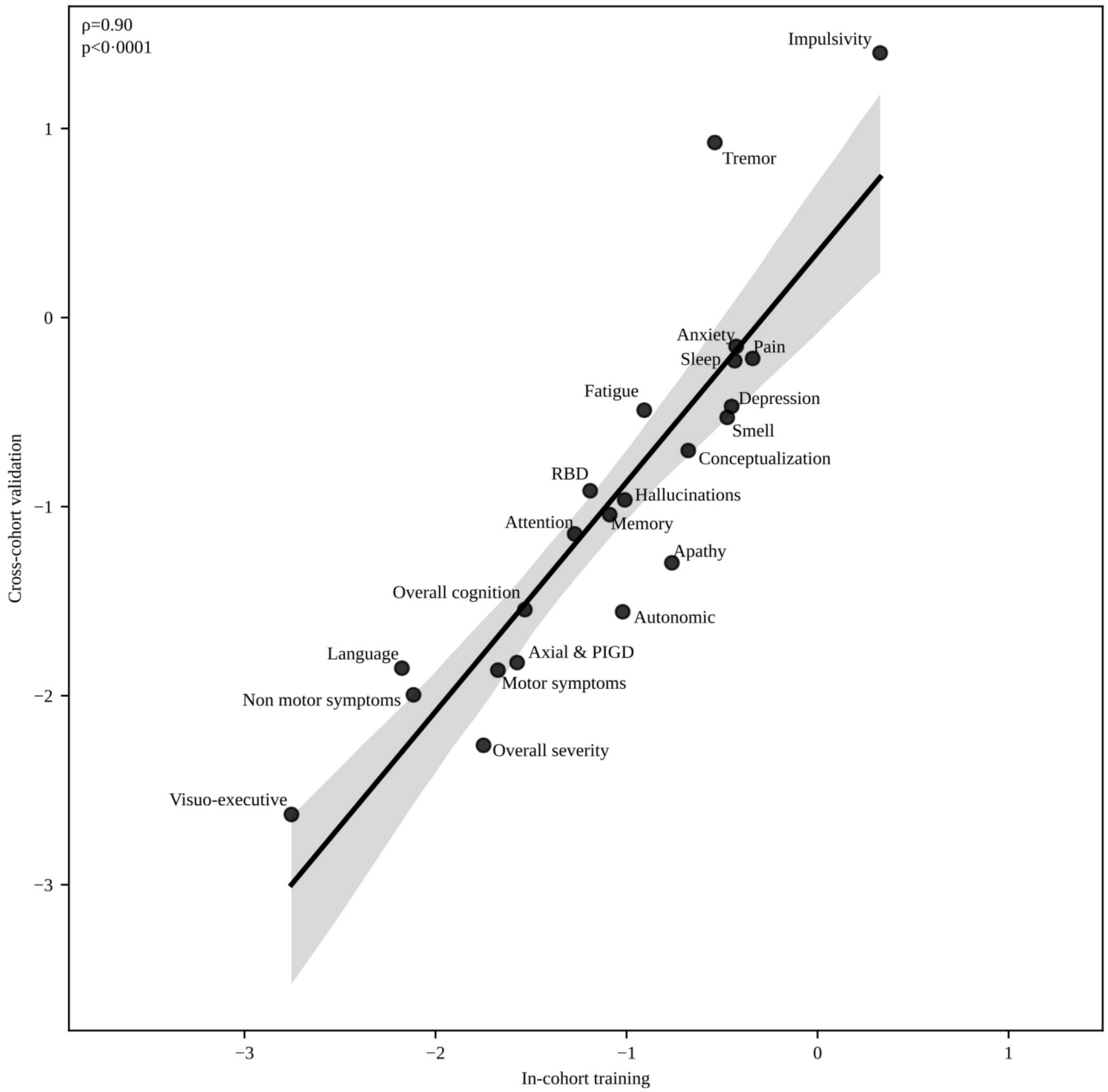
Abbreviations: BDI: Beck Depression Inventory, CERAD: Consortium to Establish a Registry for Alzheimer's Disease, CGIS: Clinical Global Impression-Severity, DAS: Dimensional Apathy Scale, ESS: Epworth Sleepiness Scale, FAB: Frontal Assessment Battery, FAQ: Functional Activities Questionnaire, FOGAC: Freezing of Gait AC, FOGQ: Freezing of Gait Questionnaire, GABS: Clinical Gait and Balance Scale, GDS: Geriatric Depression Scale, H&Y: Hoehn & Yahr scale, HADS: Hospital Anxiety and Depression Scale, MATTIS: Mattis Dementia Rating Scale, MMSE: Mini Mental Status Examination, MOCA: Montreal Cognitive Assessment, NFOGQ: New Freezing of Gait Questionnaire, NMSQ: Non-Motor Symptoms Questionnaire, PDQ39: Parkinson's Disease Questionnaire-39, PDSS: Parkinson's Disease Sleep Scale, PIGD: Postural Instability and Gait Disorder score, QUIP: Questionnaire for Impulsive-Compulsive Disorders, QUIP-RS: QUIP-Rating Scale, RBD-HK: REM Sleep Behavior Disorder Questionnaire-Hong Kong, RBD-SQ: REM Sleep Behavior Disorder Screening Questionnaire, SAS: Starkstein Apathy Scale, SCOPA-AUT: Scales for Outcomes in Parkinson's Disease-Autonomic Dysfunction, SEADL: Schwab and England Activities of Daily Living Scale, SIQCDE: Short Informant Questionnaire on Cognitive Decline in the Elderly, STA: State-Trait Anxiety Inventory, TD: Tremor Dominance Score, TUG: Timed Up and Go, UPDRS: MDS-Unified Parkinson's Disease Rating Scale, UPSIT: University of Pennsylvania Smell Identification Test, VFT: Verbal Fluency Task



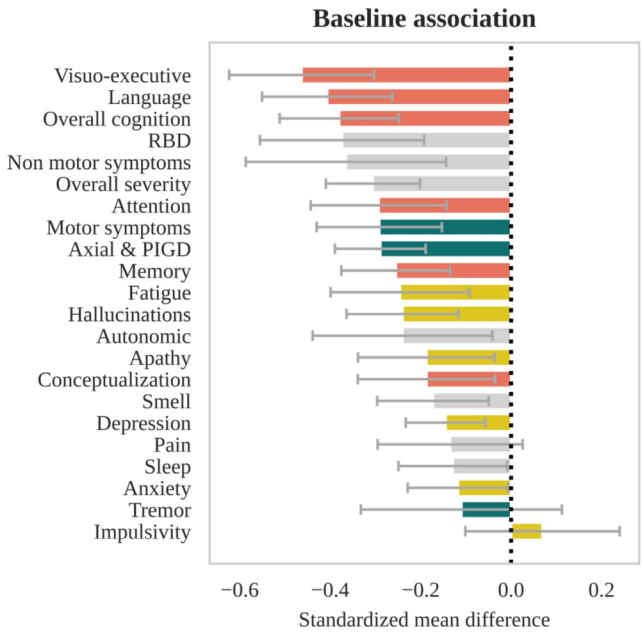
Supplementary Figure 2: Progression trajectories of subtypes for motor and non-motor symptoms (validation)
Progression of motor symptoms (UPDRS II/III/IV, PIGD) and non-motor symptoms (UPDRS I, MoCA, SCOPA) for the slow-progressing subtype (orange) and fast-progressing subtype (blue) for the ICEBERG and LuxPARK cohort. The in-cohort training results and the cross-cohort validation results (models trained on PPMI) are shown side by side. Mean and 95% confidence interval for each subtype are shown.
Abbreviations: MoCA: Montreal Cognitive Assessment, PIGD: Postural Instability and Gait Dysfunction score, SCOPA: Scales for Outcomes in Parkinson's Disease-Autonomic Dysfunction, UPDRS: Unified Parkinson's Disease Rating Scale.



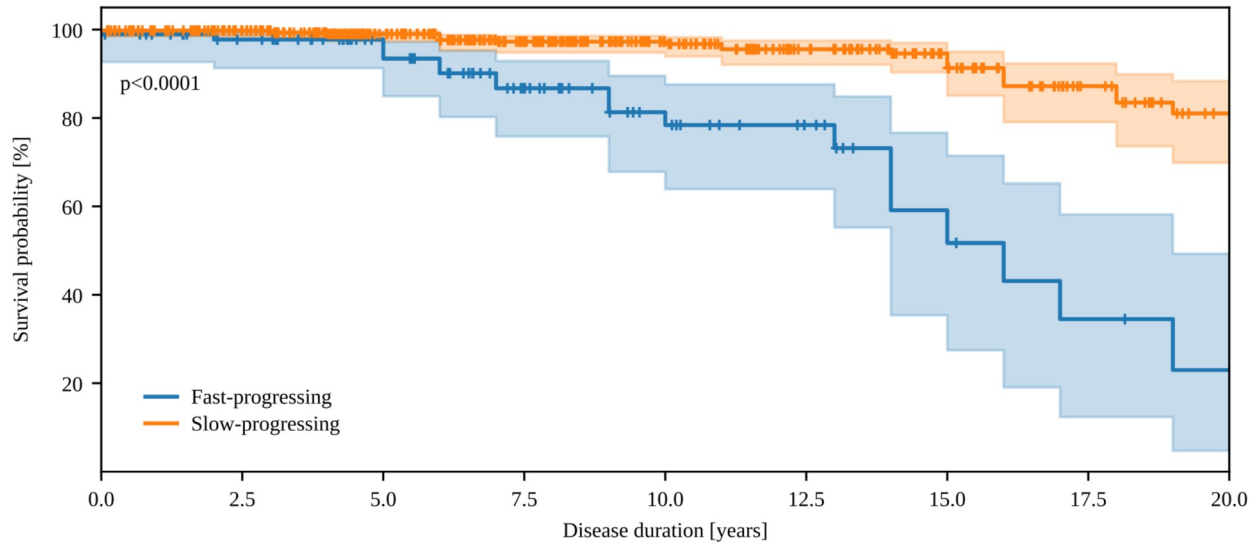
Supplementary Figure 3: Symptom domain progression rate validation
The figure depicts the correlation of standardized mean differences (SMDs) of progression rates calculated for each symptom domain using the in-cohort training approach and the cross-cohort validation approach (i.e. models trained on PPMI). The 95% confidence interval of the regression line is depicted in gray. The Pearson correlation coefficient with corresponding p-value is shown.



Supplementary Figure 4: Symptom domain baseline associations validation
The figure depicts the correlation of average regression coefficients for baseline outcomes calculated for each symptom domain using the in-cohort training approach and the cross-cohort validation approach (i.e. models trained on PPMI). The 95% confidence interval of the regression line is depicted in gray. The Pearson correlation coefficient with corresponding p-value is shown.



Supplementary Figure 5: Direct comparison of baseline characteristics between progression subtypes
Standardized mean differences (SMD) were calculated for symptom domains at baseline between subtypes. In contrast to Fig. 2C in the main manuscript, baseline outcomes were compared directly without a correction for differences in disease duration. Negative values indicate that more severe symptoms at baseline are associated with the faster subtype. 95% confidence intervals are shown and were corrected for multiple testing.
Abbreviations: PIGD: Postural Instability and Gait Dysfunction score, RBD: REM behavior sleep disorder

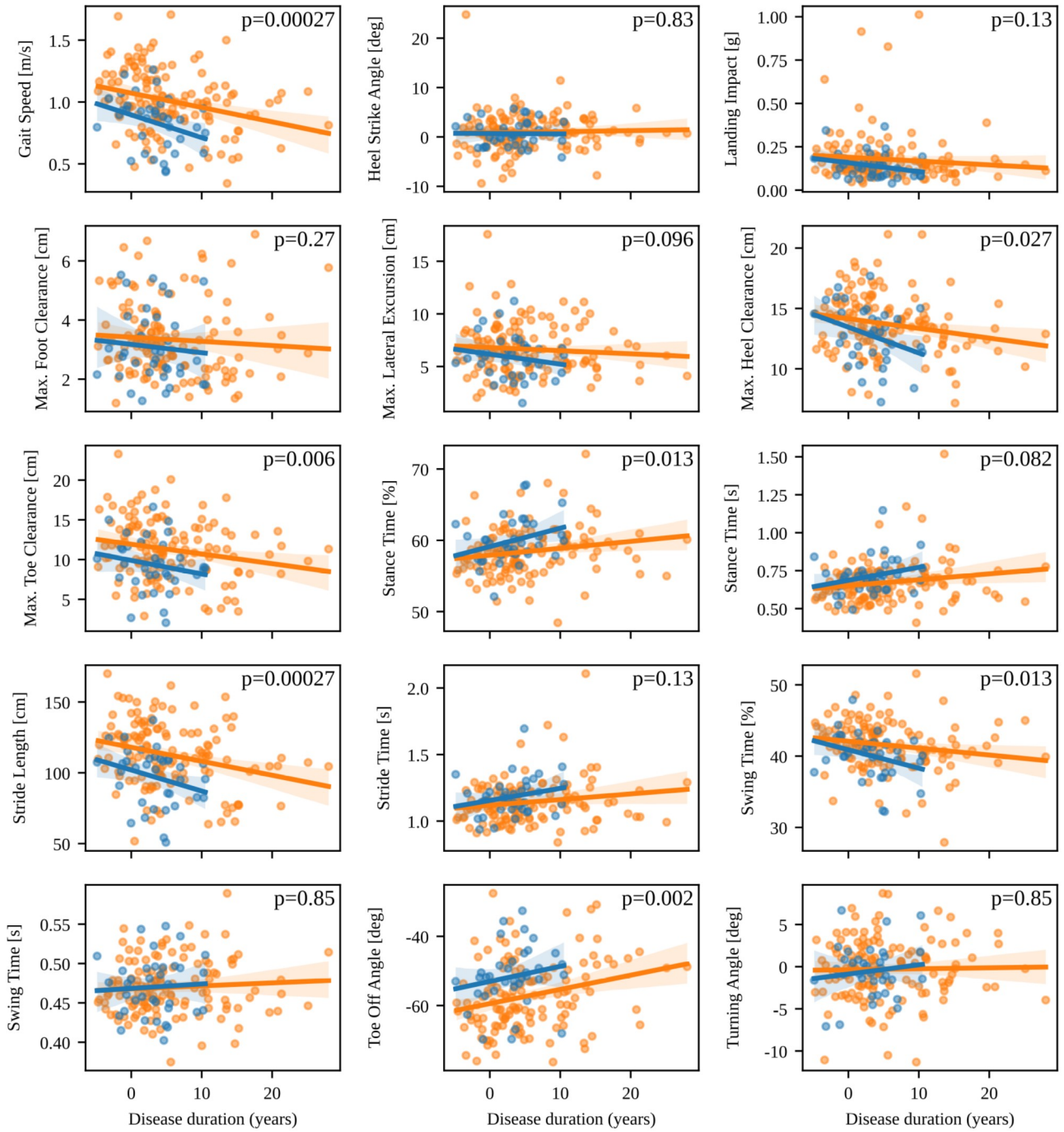


At risk	526	466	366	272	188	133	83	48	29
Censored	33	90	186	271	351	404	446	475	491
Events	2	5	9	18	22	24	32	38	41

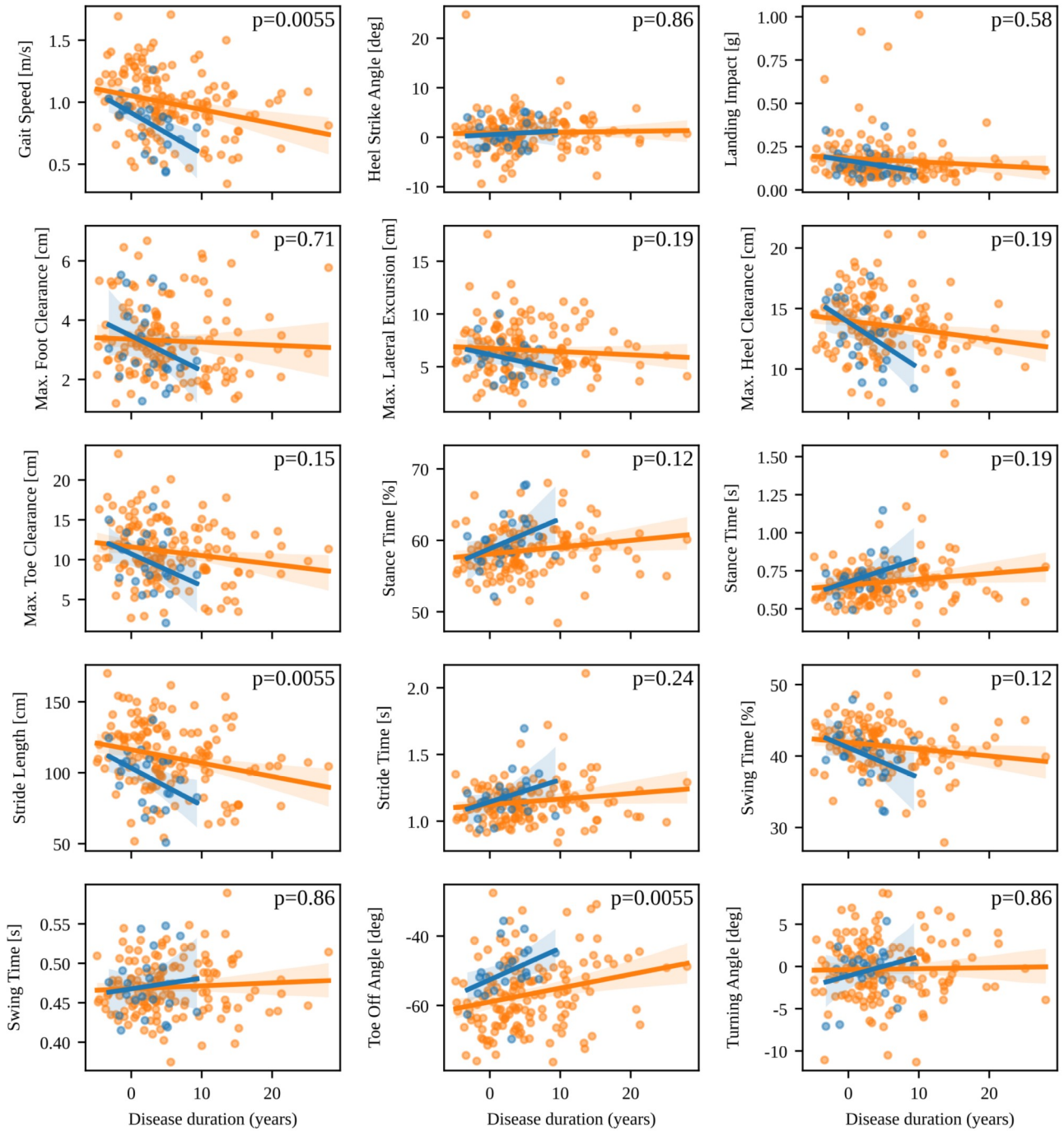
Supplementary Figure 6: Survival curves (validation)
Kaplan-Meier estimator for survival probability on the common disease timescale for fast-progressing (blue) and slow-progressing (orange) PwPD in LuxPARK. Right-censored observations are indicated by a small vertical tick. The corresponding p-value for the subtype covariate from the cox proportional hazard model is reported. 95% confidence intervals are shown. The analysis was done using the PPMI-trained model as cross-cohort validation.

Gait parameter	Unit	Description
Gait speed	m/s	The average walking speed.
Heel strike angle	degree	The angle between the toes and the surface when the foot lands.
Landing impact	g	The maximum vertical acceleration during landing of the foot.
Max. foot clearance	cm	The maximum elevation of the foot from the ground during the swing phase.
Max. heel clearance	cm	The maximum elevation of the heel from the ground during the swing phase.
Max. toe clearance	cm	The maximum elevation of the toe from the ground during the swing phase.
Max. lateral excursion	cm	The maximum lateral deviation of the foot in the swing phase, measured from an imaginary line between the foot's position at start and end of the swing phase.
Stance time (absolute)	s	Duration from initial contact of the foot with the surface until start of next swing phase of the foot.
Stance time (relative)	%	Proportion of stance time divided by the total duration of the stride.
Stride length	cm	The length of one stride.
Stride time	s	Sum of stance time and swing time.
Swing time (absolute)	s	Duration from start of swing until next foot contact with the surface.
Swing time (relative)	%	Proportion of swing time divided by the total duration of the stride.
Toe off angle	degree	The angle between the heel and the surface at the beginning of the swing phase.
Turning angle	degree	The angle between the direction of the last swing phase (imaginary line between foot position at the beginning and end of the swing phase) and the orientation of the foot in the next stance phase.

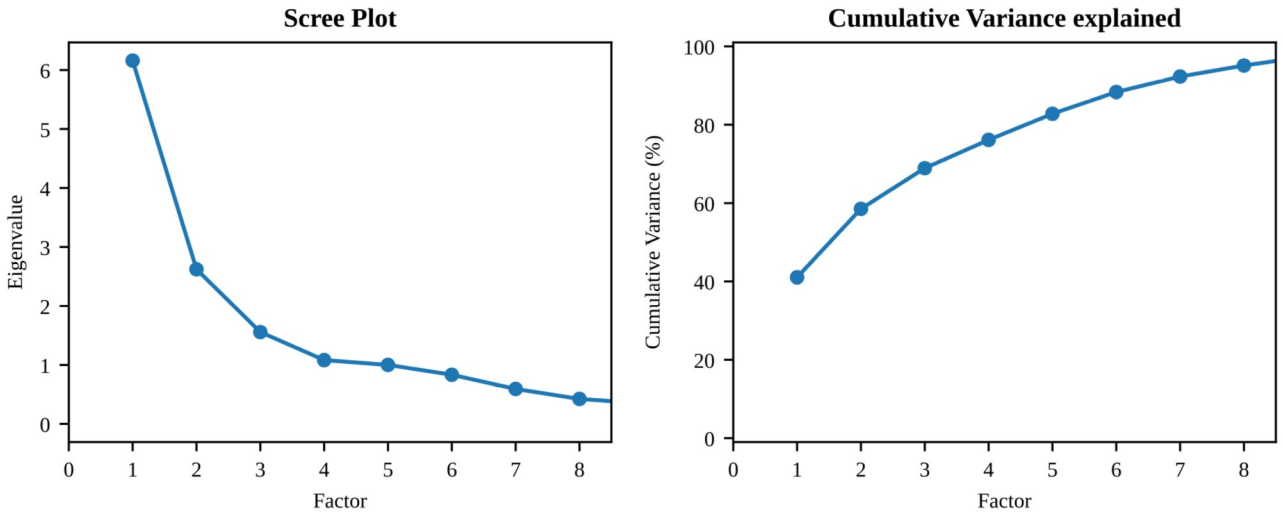
Supplementary Table 3: Description of digital gait biomarkers.
Gait parameters were calculated as mean of all straight steps from the Timed Up and Go task. Turning steps were excluded from the calculation.
Abbreviations: max: maximum, min: minimum



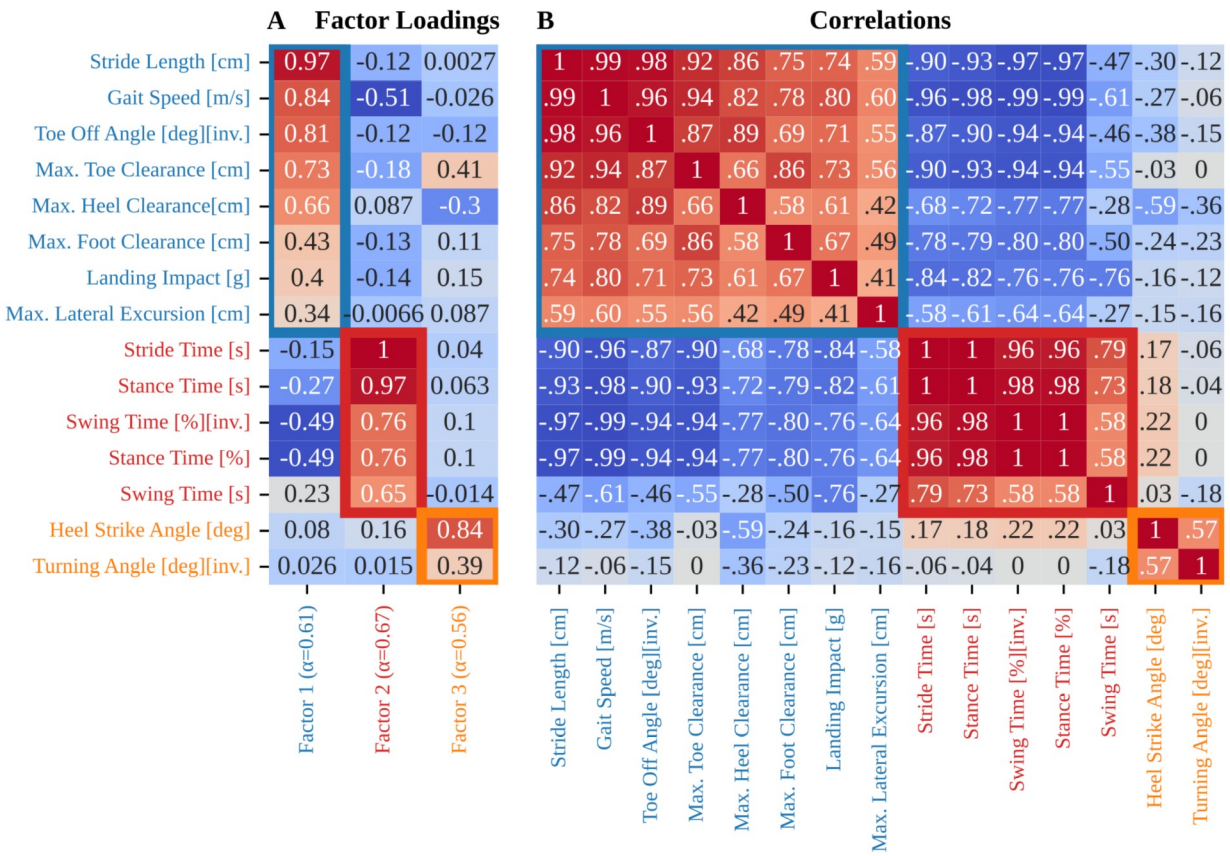
Supplementary Figure 7: Digital gait biomarkers
Correlation of all digital gait markers with disease duration on the common disease timescale for fast-progressing (blue) and slow-progressing (orange) PwPD. The corresponding p-values from the ANCOVA analyses are shown and were corrected for multiple testing. 95% confidence intervals are depicted.
Abbreviations: deg: degree.



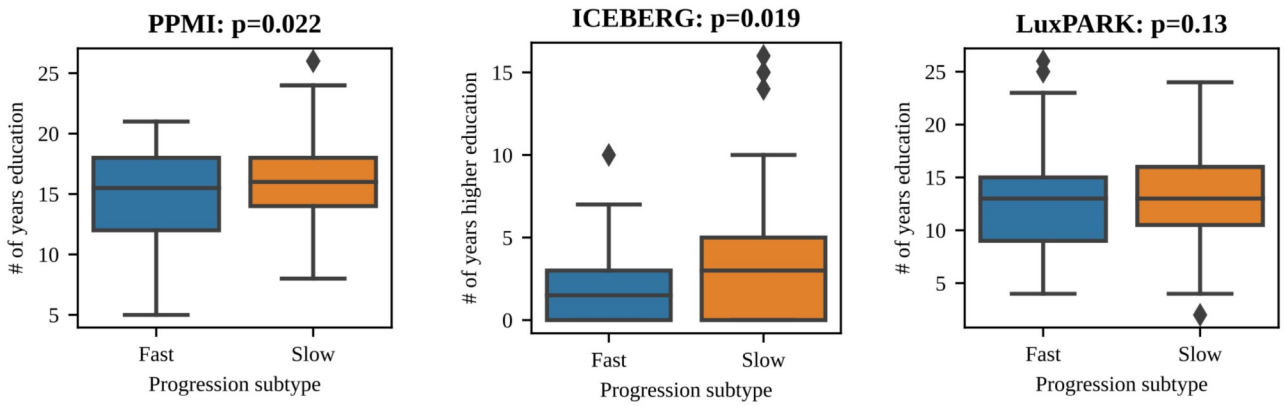
Supplementary Figure 8: Digital gait biomarkers (validation)
Correlation of all digital gait markers with disease duration on the common disease timescale for fast-progressing (blue) and slow-progressing (orange) PwPD. The corresponding p-values from the ANCOVA analyses are shown and were corrected for multiple testing. 95% confidence intervals are shown. The analysis was done using the PPMI-trained model as cross-cohort validation.
Abbreviations: deg: degree.



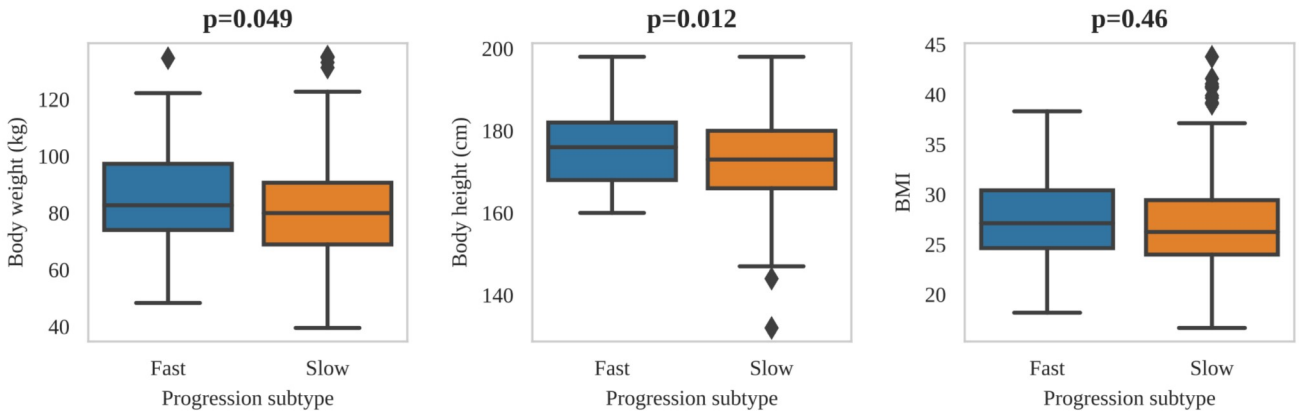
Supplementary Figure 9: Scree plot and variance explained for exploratory factor analysis of digital gait parameters
An exploratory factor analysis was performed using the LuxPARK gait data. The figure shows the Eigenvalues (left) and cumulative variance explained (right) depending on the chosen number of factors. Based on the Scree-Plot, we decided to use three factors.



Supplementary Figure 10: Factor analysis and correlation matrix for digital gait data
A*: Factor loadings for all gait features and the three factors. The choice of three factors was derived from the scree plot displayed in Supplementary Figure 9. The gait features which are mostly determined by the same factor are surrounded by a colored rectangle. The features Toe Off Angle, Swing Time, Turning Angle (marked by "[inv.]") had been inverted to enforce positive factor loadings for better visualization and for the calculation of Cronbach's alpha, which is displayed below each column. The first factor primarily encompassed the spatial sequence of the step, including measures of length and height of the step. In contrast, the second factor was based on the timing of the stance and stride phases of the steps. The third factor incorporated two angles, i.e., the heel strike angle and turning angle.* ***B****: Correlation heatmap between gait features sorted by the loadings of each factor. Gait features belonging to the same factor are surrounded by a colored rectangle.*



Supplementary Figure 11: Association of PD progression subtypes with educational level
Comparison of education levels between both progression subtypes, i.e., number of education years (PPMI, LuxPARK) and number of years with higher education (ICEBERG). Correction for multiple testing was performed using Benjamini-Hochberg procedure. The boxplots are displayed with a median line, box borders representing the interquartile range (IQR), whiskers extending to 1.5 times the IQR, and outliers depicted as diamonds beyond the whiskers.



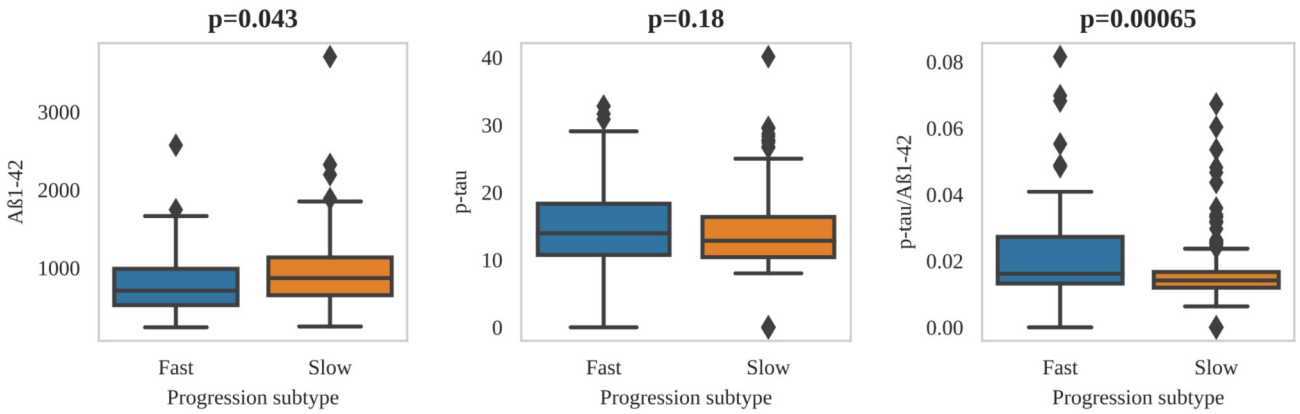
Supplementary Figure 12: Association of PD progression subtypes with body weight, body height and body mass index
All analyzes were performed for PPMI. P-values were corrected for age and sex. Correction for multiple testing was performed using Benjamini-Hochberg procedure. The boxplots are displayed with a median line, box borders representing the interquartile range (IQR), whiskers extending to 1.5 times the IQR, and outliers depicted as diamonds beyond the whiskers.

Blood pressure parameter	p-value (uncorrected)	p-value (corrected for multiple testing)
Systolic blood pressure (supine)	0.45	0.68
Diastolic blood pressure (supine)	0.62	0.74
Systolic blood pressure (standing)	0.44	0.68
Diastolic blood pressure (standing)	0.24	0.68
Systolic blood pressure drop (supine - standing)	0.96	0.96
Diastolic blood pressure drop (supine - standing)	0.38	0.68

Supplementary Table 4: Association of PD progression subtypes with blood pressure
All analyzes were performed for PPMI. P-values were corrected for age and sex. Correction for multiple testing was performed using Benjamini-Hochberg procedure.

Comorbidity group	p-value (uncorrected)	p-value (corrected for multiple testing)
Pulmonary	*0.033*	0.26
Ophthalmological	*0.035*	0.26
Metabolic/Endocrine	0.16	0.81
Dermatological	0.26	0.87
Psychiatric	0.46	0.87
Cardiovascular	0.47	0.87
Musculoskeletal	0.48	0.87
Allergy/immunologic	0.56	0.87
Other	0.67	0.87
Gastrointestinal	0.68	0.87
Gynecological/Urologic	0.71	0.87
Renal	0.75	0.87
Hemato/Lymphatic	0.75	0.87
ENT	0.86	0.92
Hepatobiliary	0.96	0.96

Supplementary Table 5: Association of PD progression subtypes with diagnoses groups
All analyzes were performed for PPMI. P-values were corrected for age and sex. Correction for multiple testing was performed using Benjamini-Hochberg procedure.

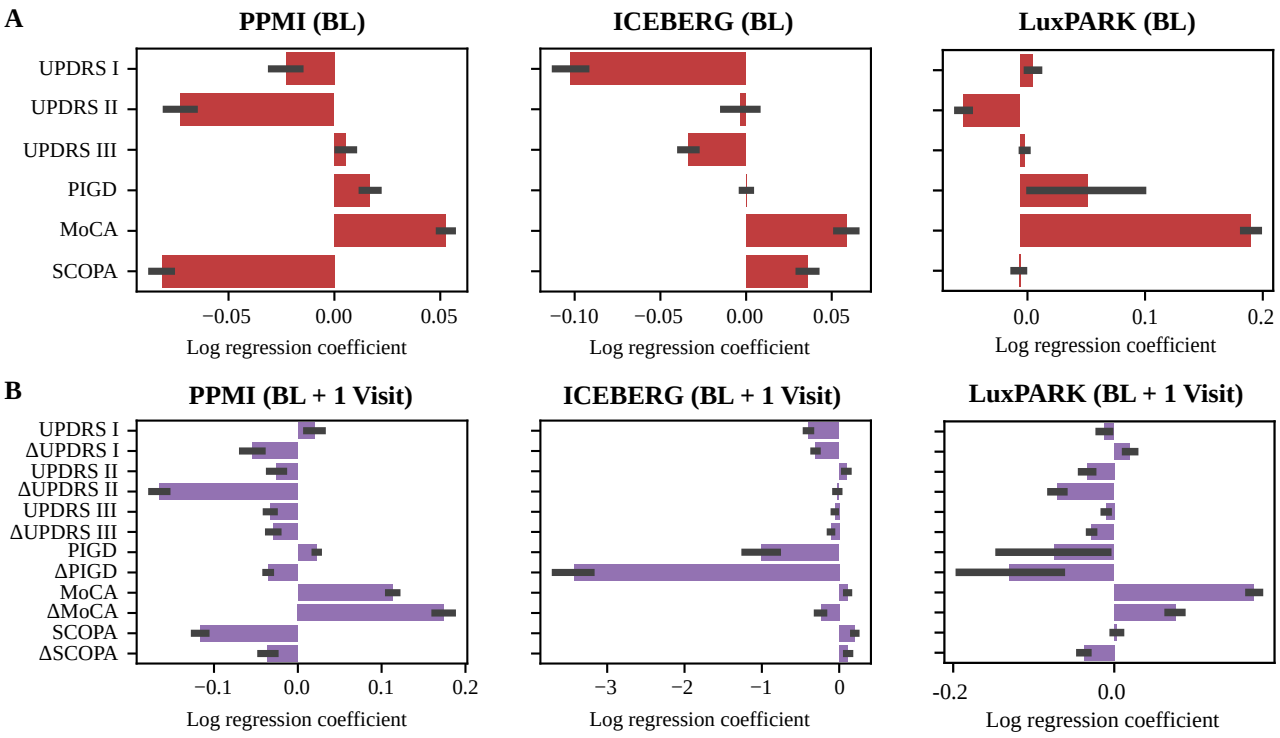


Supplementary Figure 13: Association of PD progression subtypes with Alzheimer's disease pathology cerebrospinal fluid biomarkers

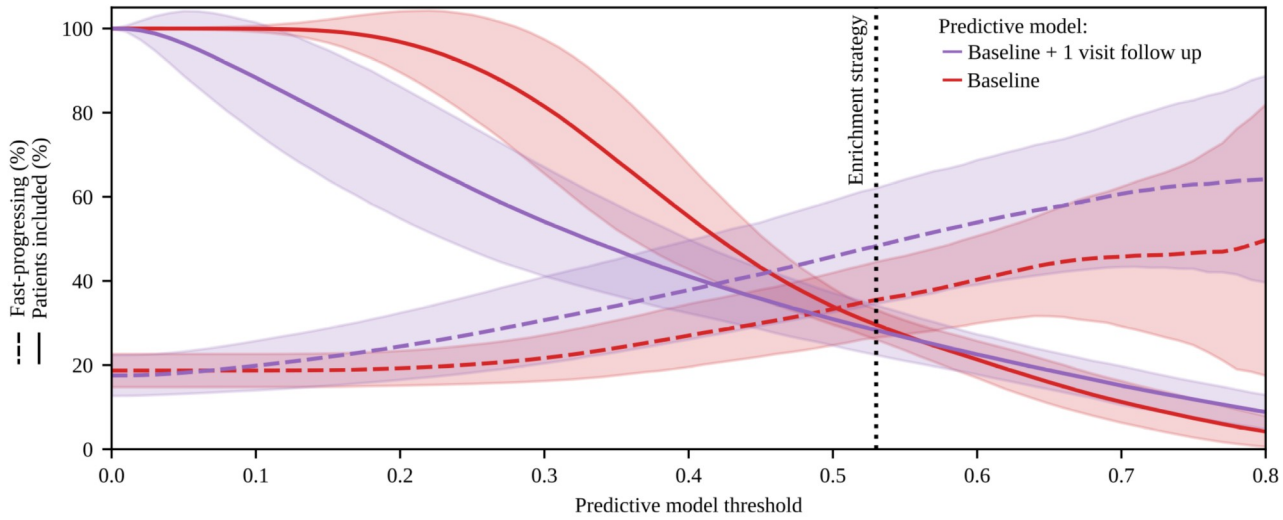
All analyzes were performed for PPMI. P-values were corrected for age and sex. Correction for multiple testing was performed using Benjamini-Hochberg procedure. The boxplots are displayed with a median line, box borders representing the interquartile range (IQR), whiskers extending to 1.5 times the IQR, and outliers depicted as diamonds beyond the whiskers.

Comedication group	p-value (uncorrected)	p-value (corrected for multiple testing)
Calcium Channel Blockers	0.18	0.95
Beta Antagonists	0.31	0.95
NSARs (without ASS)	0.44	0.95
NSAR	0.53	0.95
Ibuprofen	0.63	0.95
Statins	0.73	0.95
ASS	0.83	0.95
Contraceptives	1.0	1.0

Supplementary Table 6: Association of PD progression subtypes with specific medications.
All analyzes were performed for PPMI. P-values were corrected for age and sex. Correction for multiple testing was performed using Benjamini-Hochberg procedure.



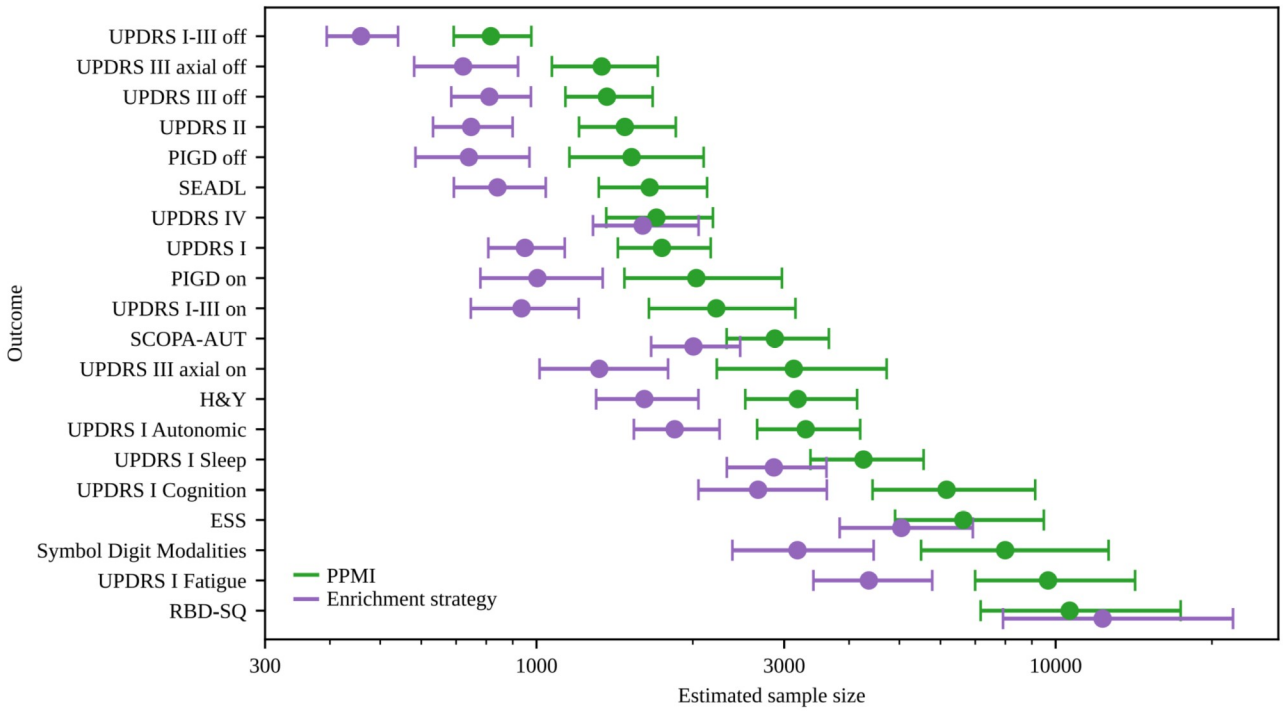
Supplementary Figure 14: Coefficients of logistic regression models for subtype predictions
*Logistic regression coefficients with 95% confidence interval using baseline data (**A**, red) or baseline data with one follow up visit (**B**, purple) from the predictive logistic regression model are shown. Higher outcome scores together with positive coefficients influence the prediction towards the slow-progressing subtype and towards the fast-progressing subtype if coefficients are negative.*
Abbreviations: BL: Baseline, Δ: Difference from first visit to baseline, MoCA: Montreal Cognitive Assessment, PIGD: Postural Instability and Gait Dysfunction score, SCOPA: Scales for Outcomes in Parkinson's Disease-Autonomic Dysfunction, UPDRS: Unified Parkinson's Disease Rating Scale.



Supplementary Figure 15: Trade-off between enriching fast-progressing PwPD and number of eligible PwPD
Depending on the chosen threshold applied to the logistic predictive model, the percentage of fast-progressing PwPD (dashed lines) and the percentage of PwPD being still eligible for study inclusion (solid lines) changes. Curves are displayed for the predictive model using baseline data (red) and the predictive model using data from baseline and one follow-up visit (purple). The vertical dotted line indicates the threshold of the predictive model, where still 30% of PwPD are eligible for study inclusion and a 47% enrichment of fast-progressing PwPD is achieved using the predictive model based on baseline data and one follow-up visit. 95% confidence intervals are depicted.

% fast-progressing PwPD	PPMI	ICEBERG	LuxPARK
No enrichment	18 %	27 %	26 %
BL enrichment (in-cohort)	36 %	43 %	47 %
BL+FU enrichment (in-cohort)	47 %	65 %	53 %
BL enrichment (cross-cohort)	-	27 %	38 %
BL+FU enrichment (cross-cohort)	-	38 %	41 %

Supplementary Table 7:Enrichment of fast-progressing PwPD using different predictive models
Predictive models were based on baseline data (BL) or baseline with one follow up visit (BL+FU). Models were trained on the same data set (in-cohort), or PPMI and cross-validated on ICEBERG/LuxPARK (cross-cohort). Using the predictive models and an appropriate threshold for the three cohorts (PPMI, ICEBERG, LuxPARK), the reported fractions of fast-progressing PwPD can be achieved in a study cohort by still allowing inclusion of 30% of all PwPD. The *"No enrichment" row reports the fraction of fast-progressing in the original cohorts without applying any enrichment strategy.*
Abbreviations: BL: baseline, FU: follow up



Supplementary Figure 16: Subtype enrichment for sample size reduction in clinical trials using different outcomes
Required sample sizes for a clinical trial depending on the clinical score used as primary outcome in the trial. The required sample sizes calculated using the default PPMI cohort are shown in green. The required sample sizes using the enrichment strategy from Fig. 5A are shown in purple. Mean estimate and 95% confidence intervals are shown. Sample sizes were calculated for all outcomes listed in Supplementary Table 2 and assessed in PPMI. From these outcomes, only the 20 outcomes with lowest required sample sizes are shown. The estimated sample size is shown on a logarithmic axis.
Abbreviations: ESS: Epworth Sleepiness Scale, H&Y: Hoehn & Yahr, PIGD: Postural Instability and Gait Disorder score, RBD-SQ: REM Sleep Behavior Disorder Screening Questionnaire, SCOPA-AUT: Scales for Outcomes in Parkinson's Disease-Autonomic Dysfunction, SEADL: Schwab and England Activities of Daily Living Scale, UPDRS: Unified Parkinson's Disease Rating Scale.

% sample size reduction	PPMI	ICEBERG	LuxPARK
BL (in-cohort)	30 %	32 %	44 %
BL+FU (in-cohort)	43 %	56 %	52 %
BL (cross-cohort)	-	0 %	28 %
BL+FU (cross-cohort)	-	36 %	34 %

Supplementary Table 8: Sample size reduction in a simulated clinical trial using different predictive models for enrichment of fast-progressing PwPD

Predictive models were based on baseline data (BL) or baseline with one follow up visit (BL+FU). Models were trained on the same data set (in-cohort) or PPMI and cross-validated on ICEBERG/LuxPARK (cross-cohort). Using the predictive models and an appropriate threshold to the three cohorts (PPMI, ICEBERG, LuxPARK), the reported sample size reductions can be achieved by still allowing inclusion of 30% of all PwPD.

Abbreviations: BL: baseline, FU: follow up

	PPMI	ICEBERG	LuxPARK
number of clusters	2	2	2
learning rate	0.0001	0.001	0.0001
batch size	16	32	16
number of nodes (first hidden layer)	64	32	32
number of nodes (second hidden layer)	1	1	2

Supplementary Table 9: Results of VaDER hyperparameter optimization for PPMI, ICEBERG and LuxPARK

Model	**Parameter**	**Grid values**	**Baseline** **PPMI / ICEBERG / LuxPARK**	**Baseline + FU** **PPMI / ICEBERG / LuxPARK**
Logistic regression	Lambda	0.001, 0.01, 0.1, 1, 10, 100, 1000	0.01 / 0.01 / 100	0.01 / 10 / 10
Random Forest	Bootstrap	True, False	True / True / True	True / True / True
	Maximum depth	10, 20, 30, 40	40 / 20 / 20	20 / 10 / 10
	Minimum samples per leaf	1, 2, 4	4 / 1 / 1	4 / 2 / 1
	Minimum samples per split	2, 5, 10	10 / 5 / 10	10 / 10 / 10
	Number of Estimators	200, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000	800 / 1400 / 1800	1800 / 2000 / 200
XGBoost	Minimum Child Weight	1, 5, 10	10 / 1 / 1	5 / 1 / 5
	Gamma	0.5, 1, 1.5, 2, 5	2 / 0.5 / 5	0.5 / 1.5 / 1.5
	Subsample	0.6, 0.8, 1.0	0.8 / 0.6 / 0.8	0.8 / 0.6 / 1.0
	Colsample by tree	0.6, 0.8, 1.0	0.8 / 0.6 / 0.6	0.6 / 1.0 / 1.0
	Maximum Depth	3, 4, 5	4 / 4 / 3	4 / 3 / 3

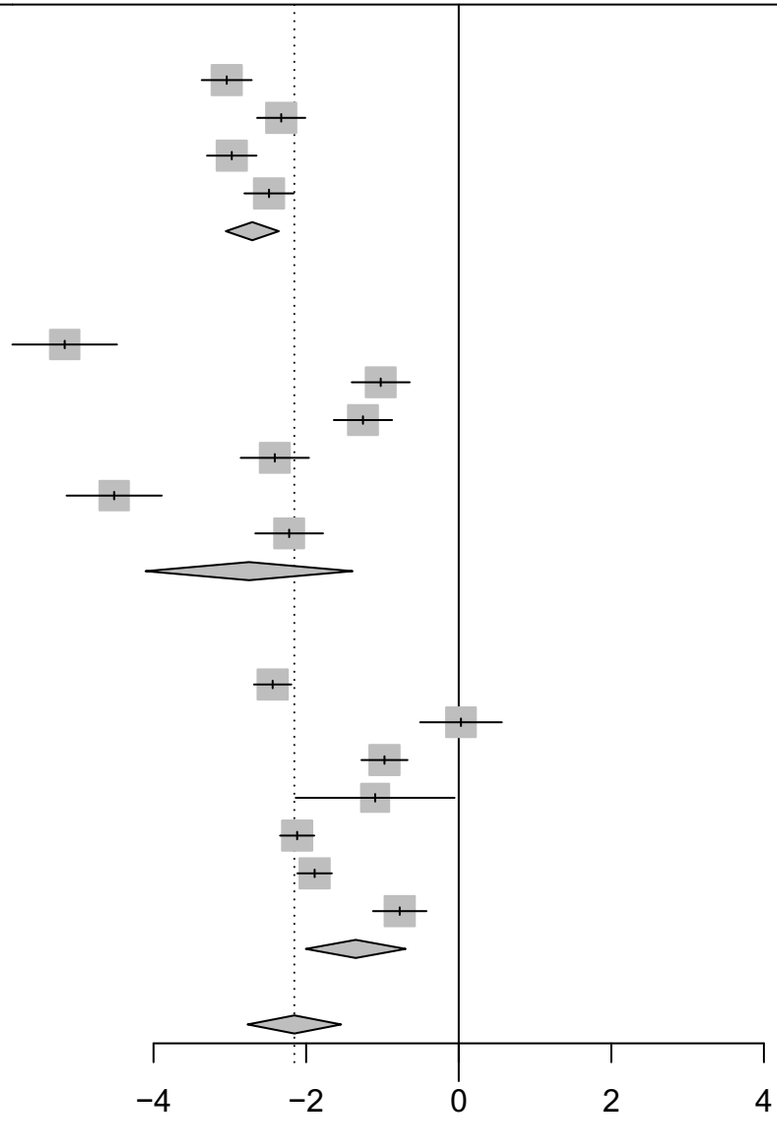
Supplementary Table 10: Predictive model hyperparameter optimization
Hyperparameter space used for hyperparameter optimization. For Logistic regression, L2 penalization and hyperparameter grid search was used. For Random Forest and XGBoost, randomized hyperparameter search with 50 samples was used. Repeated stratified k-fold cross validation was performed using 5 splits and 20 repeats. Results of hyperparameter training for the baseline-only and baseline with one visit follow-up models are presented.
Abbreviations: FU: follow up

Progression subtypes in Parkinson’s disease identified by a data driven multi cohort analysis

Forest plots for symptom domain progression (in cohort)

Forest plot for progression characteristics of symptom domain Axial & PIGD

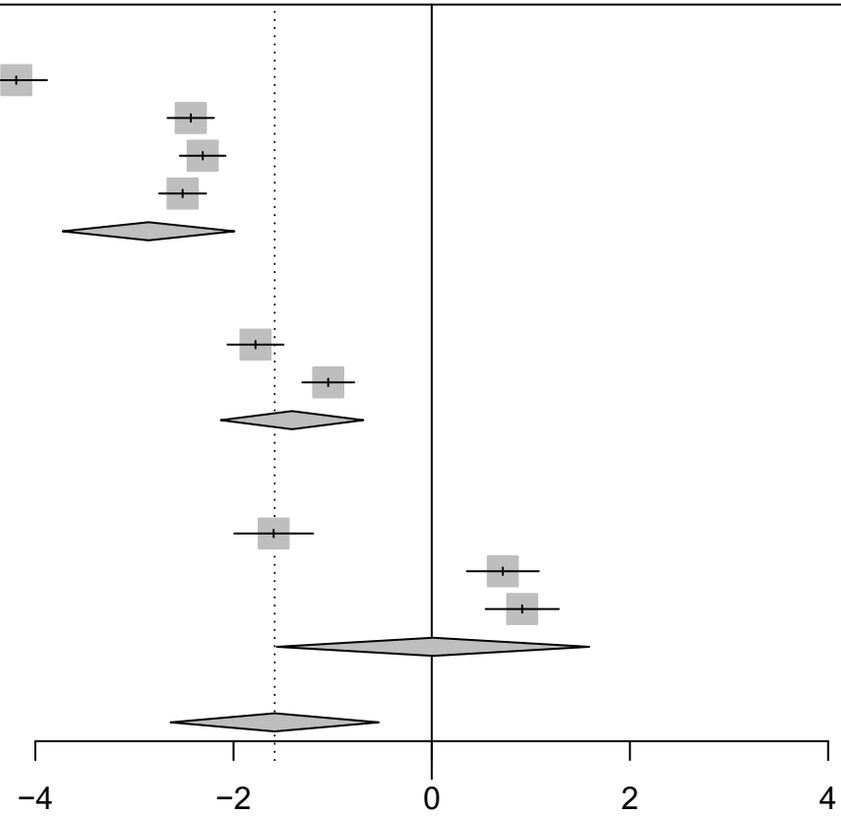
Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = PPMI								
UPDRS III axial off	335	0.16	0.0777	74	0.52	0.2251	−3.04	[−3.37; −2.71]
UPDRS III axial on	300	0.09	0.0855	66	0.43	0.2922	−2.33	[−2.64; −2.01]
PIGD off	335	0.13	0.0917	74	0.61	0.3265	−2.97	[−3.30; −2.65]
PIGD on	300	0.10	0.0823	66	0.48	0.3223	−2.49	[−2.81; −2.16]
Random effects model	**1270**			**280**			**−2.71**	**[−3.05; −2.36]**
Cohort = ICEBERG								
UPDRS III axial off	111	0.09	0.0004	42	0.36	0.1002	−5.16	[−5.85; −4.48]
UPDRS III axial on	105	0.10	0.0011	41	0.17	0.1309	−1.02	[−1.40; −0.64]
GABS Examination	112	0.04	0.0191	42	0.12	0.1273	−1.26	[−1.64; −0.87]
GABS Questionaire	112	0.03	0.0138	42	0.15	0.0932	−2.41	[−2.86; −1.96]
PIGD off	111	0.06	0.0154	42	0.24	0.0735	−4.52	[−5.14; −3.89]
PIGD on	105	0.07	0.0042	41	0.23	0.1364	−2.22	[−2.67; −1.78]
Random effects model	**656**			**250**			**−2.75**	**[−4.10; −1.39]**
Cohort = LuxPARK								
UPDRS III axial on	380	0.13	0.0510	140	0.30	0.1008	−2.44	[−2.68; −2.19]
FOGAC	63	0.11	0.0767	17	0.11	0.1963	0.03	[−0.51; 0.57]
FOGQ	166	0.17	0.0954	64	0.31	0.2075	−0.97	[−1.28; −0.67]
NFOGQ	19	0.24	0.0939	5	0.91	1.3814	−1.10	[−2.14; −0.05]
PDQ39 Mobility	407	0.18	0.0835	152	0.44	0.1897	−2.12	[−2.34; −1.89]
PIGD on	380	0.14	0.0843	140	0.33	0.1427	−1.89	[−2.11; −1.66]
TuG	154	0.03	0.0318	42	0.05	0.0269	−0.77	[−1.12; −0.42]
Random effects model	**1569**			**560**			**−1.35**	**[−2.00; −0.70]**
Random effects model	**3495**			**1090**			**−2.15**	**[−2.77; −1.54]**



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Axial & PIGD across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Depression

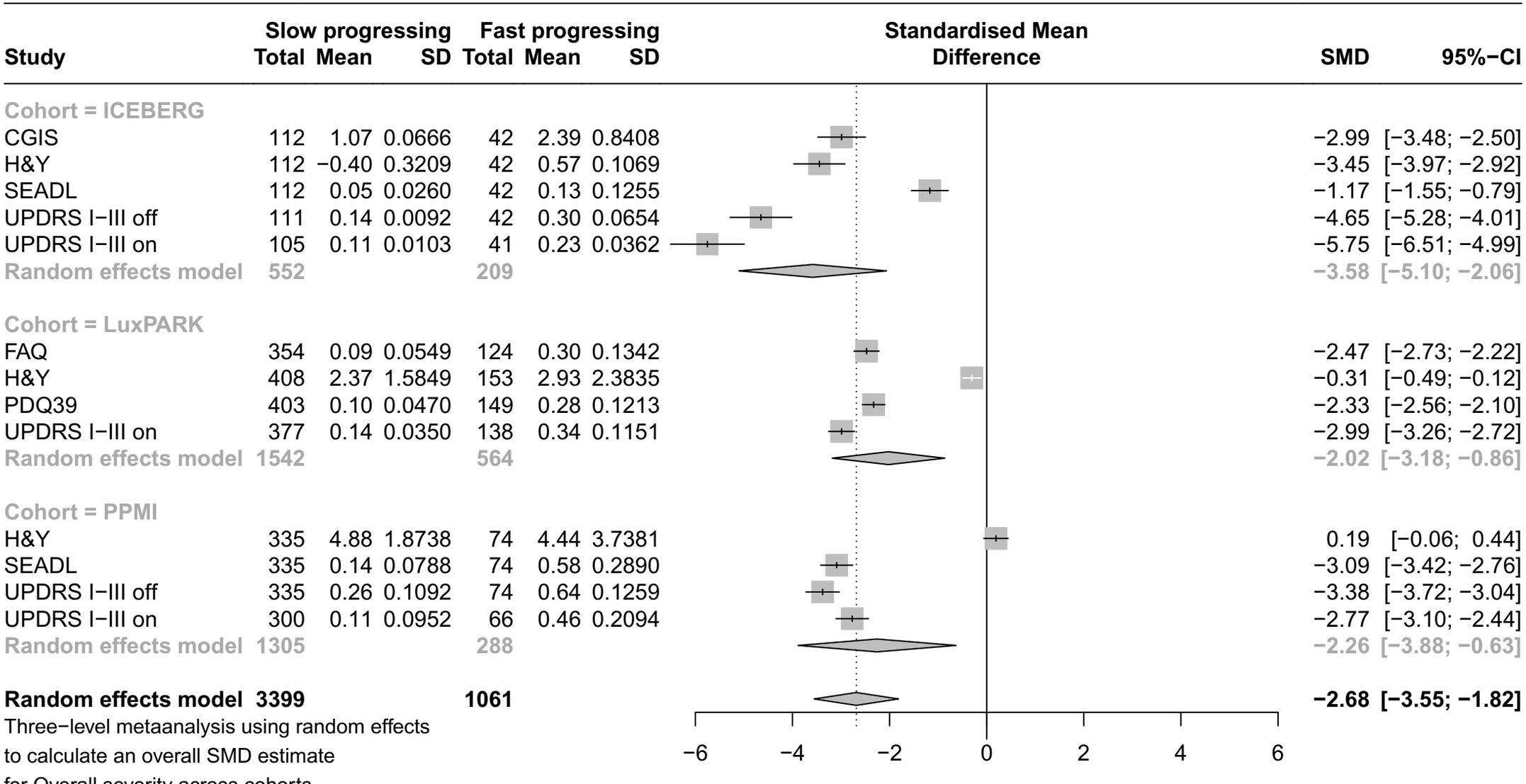
Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = LuxPARK								
BDI	406	0.04	0.0176	149	0.14	0.0358	−4.19	[−4.50; −3.88]
NMSQ Depression	408	1.04	0.5080	152	2.15	0.2609	−2.43	[−2.67; −2.20]
PDQ39 Depression	407	0.06	0.0499	152	0.18	0.0605	−2.31	[−2.54; −2.08]
UPDRS I Depression	408	0.84	0.1665	153	1.53	0.4505	−2.51	[−2.75; −2.28]
Random effects model	**1629**			**606**			**−2.86**	**[−3.72; −1.99]**
Cohort = PPMI								
GDS	324	0.05	0.1681	73	0.35	0.1441	−1.78	[−2.06; −1.50]
UPDRS I Depression	335	1.16	1.4484	74	2.83	2.1275	−1.04	[−1.31; −0.78]
Random effects model	**659**			**147**			**−1.41**	**[−2.13; −0.69]**
Cohort = ICEBERG								
HADS depression	112	0.10	0.0186	42	0.16	0.0578	−1.60	[−1.99; −1.20]
NMSQ Depression	112	2.00	0.4071	42	1.42	1.4046	0.72	[0.35; 1.08]
UPDRS I Depression	112	1.54	0.5269	42	0.83	1.2172	0.91	[0.54; 1.28]
Random effects model	**336**			**126**			**0.01**	**[−1.56; 1.59]**
Random effects model	**2624**			**879**			**−1.58**	**[−2.64; −0.53]**



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Depression across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

<− Faster progression for fast−progressing subtype | faster progression for slow−progressing subtype −>

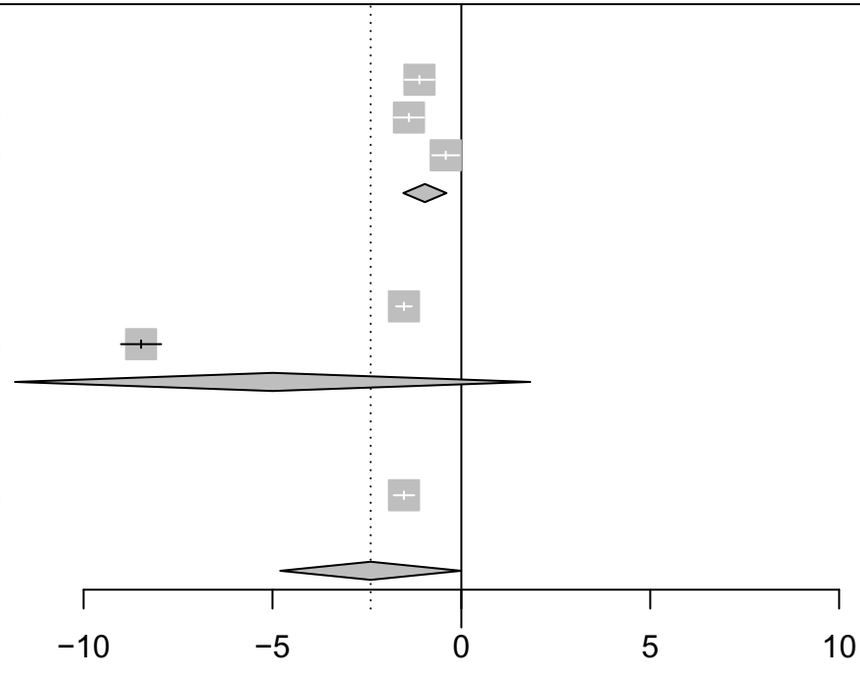
Forest plot for progression characteristics of symptom domain Overall severity



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Overall severity across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Apathy

Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = ICEBERG								
DAS	103	0.08	0.0179	38	0.11	0.0428	−1.11	[−1.50; −0.71]
SAS	104	0.08	0.0303	41	0.12	0.0322	−1.39	[−1.78; −0.99]
UPDRS I Apathy	112	1.18	0.4818	42	1.44	0.8770	−0.41	[−0.77; −0.06]
Random effects model	**319**			**121**			**−0.96**	**[−1.54; −0.39]**
Cohort = LuxPARK								
SAS	405	0.05	0.0480	153	0.12	0.0369	−1.52	[−1.73; −1.31]
UPDRS I Apathy	408	0.73	0.0431	153	1.86	0.2459	−8.48	[−9.01; −7.95]
Random effects model	**813**			**306**			**−4.99**	**[−11.81; 1.82]**
Cohort = PPMI								
UPDRS I Apathy	335	2.14	1.2773	74	4.63	2.7266	−1.52	[−1.79; −1.25]
Random effects model	**1467**			**501**			**−2.40**	**[−4.80; −0.00]**

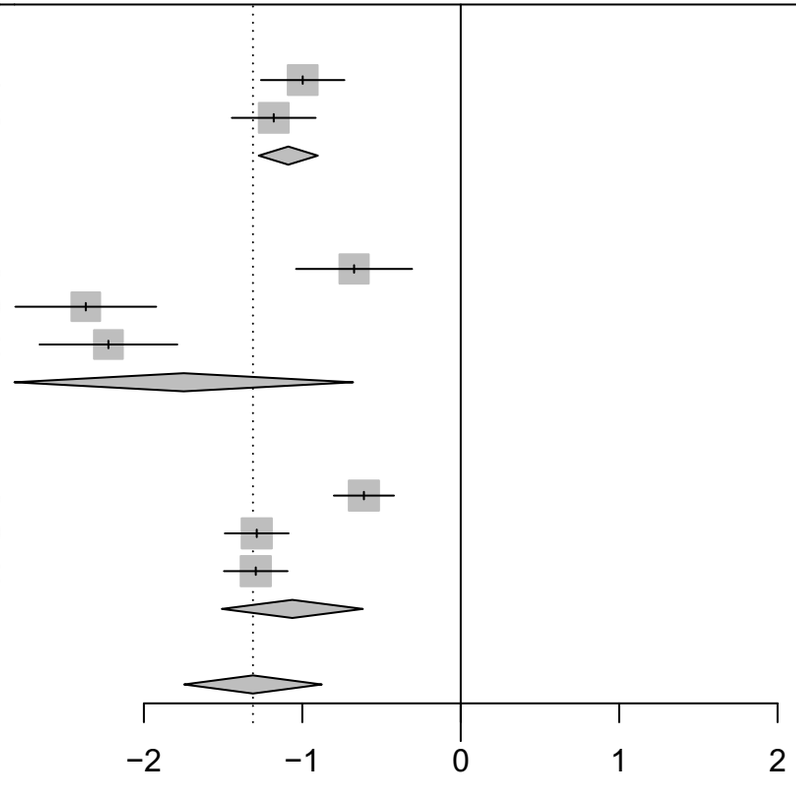


Three−level metaanalysis using random effects to calculate an overall SMD estimate for Apathy across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

<− Faster progression for fast−progressing subtype | faster progression for slow−progressing subtype −>

Forest plot for progression characteristics of symptom domain Sleep

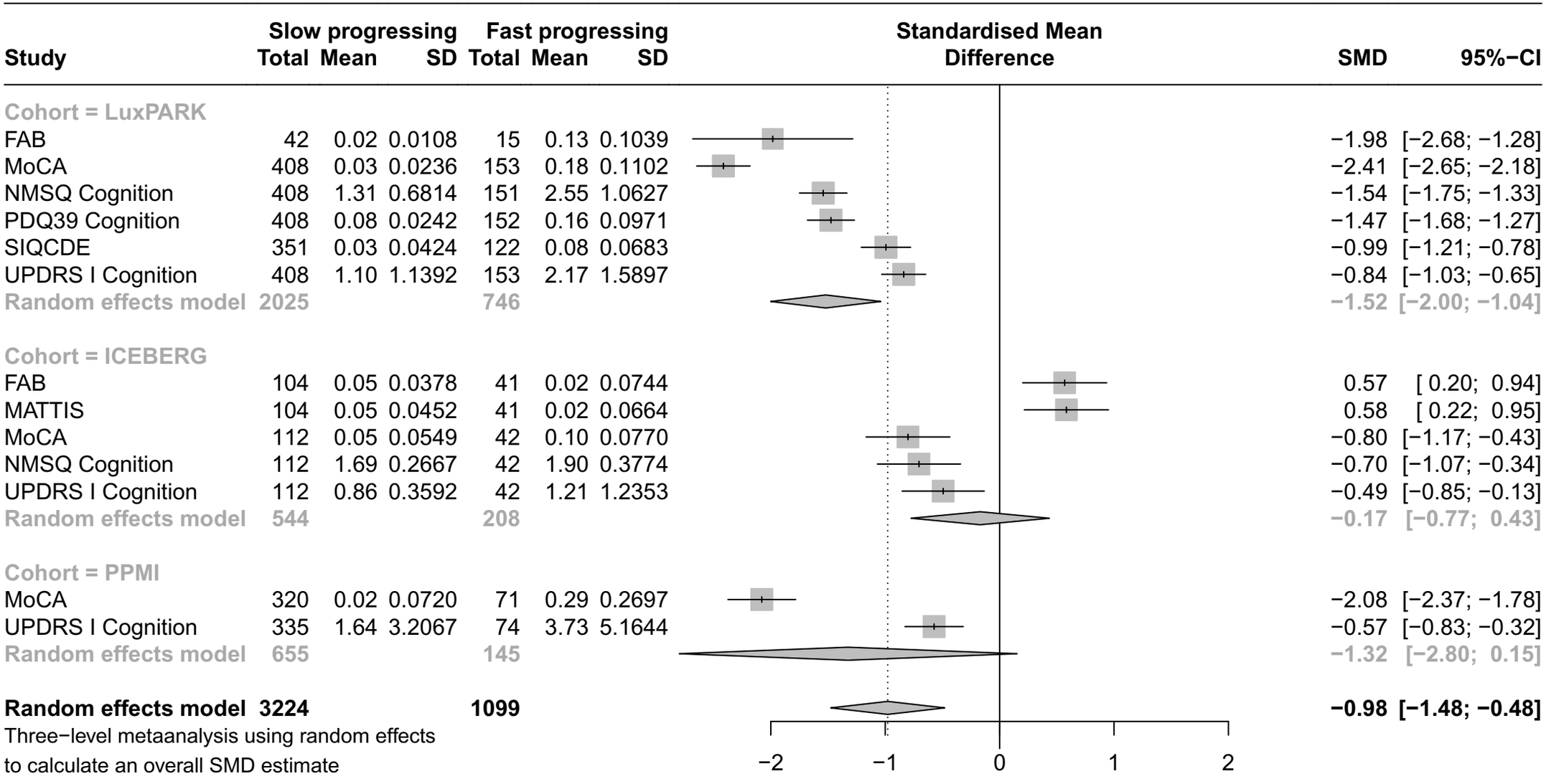
Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = PPMI								
ESS	324	0.15	0.1390	73	0.28	0.1024	−1.00	[−1.26; −0.73]
UPDRS I Sleep	335	0.25	0.1857	74	0.48	0.2186	−1.18	[−1.44; −0.92]
Random effects model	659			147			−1.09	[−1.28; −0.90]
Cohort = ICEBERG								
ESS	111	0.15	0.1219	41	0.22	0.0565	−0.67	[−1.04; −0.31]
NMSQ Sleep	112	1.08	0.5661	42	2.25	0.1739	−2.37	[−2.81; −1.92]
UPDRS I Sleep	112	0.07	0.0314	42	0.13	0.0082	−2.22	[−2.66; −1.79]
Random effects model	335			125			−1.75	[−2.82; −0.68]
Cohort = LuxPARK								
NMSQ Sleep	408	0.64	0.0903	152	0.83	0.5505	−0.61	[−0.80; −0.42]
PDSS	406	0.10	0.0189	152	0.13	0.0379	−1.29	[−1.49; −1.09]
UPDRS I Sleep	408	0.09	0.0213	153	0.13	0.0592	−1.29	[−1.49; −1.09]
Random effects model	1222			457			−1.06	[−1.51; −0.62]
Random effects model	**2216**			**729**			**−1.31**	**[−1.75; −0.88]**



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Sleep across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

<− Faster progression for fast−progressing subtype | faster progression for slow−progressing subtype −>

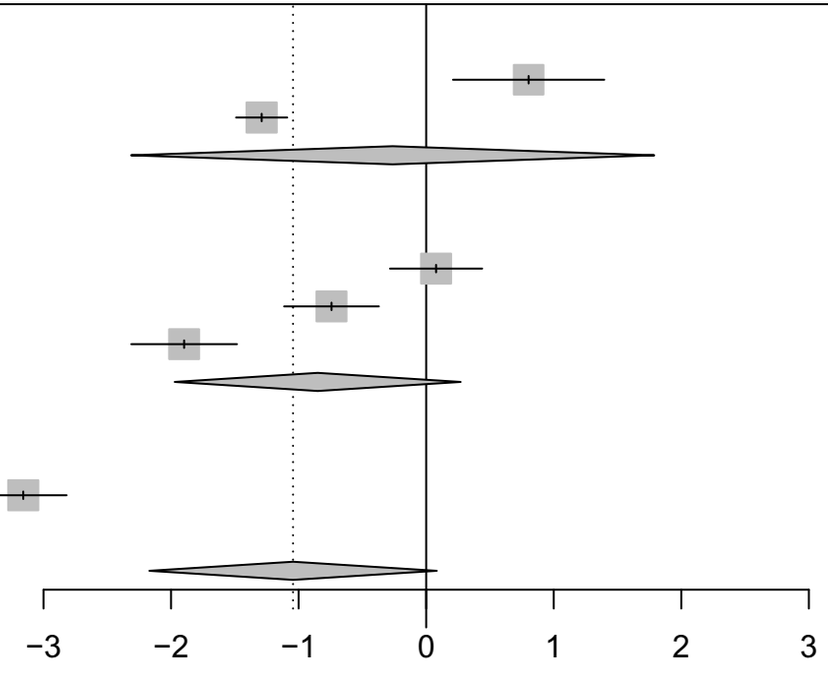
Forest plot for progression characteristics of symptom domain Overall cognition



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Overall cognition across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Conceptualization

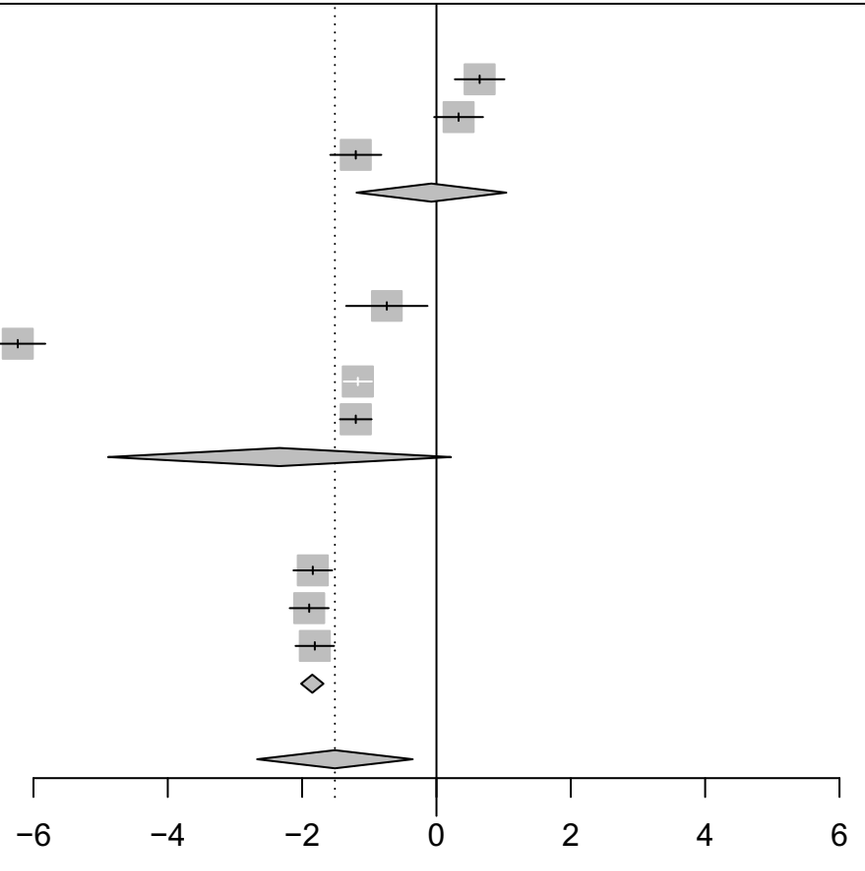
Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = LuxPARK								
FAB Conceptualization	43	−1.31	0.2084	16	−2.15	1.9839	0.80	[0.21; 1.40]
MOCA Abstraction	408	0.16	0.2392	153	0.43	0.1078	−1.29	[−1.49; −1.09]
Random effects model	**451**			**169**			**−0.26**	**[−2.31; 1.79]**
Cohort = ICEBERG								
FAB Conceptualization	104	0.97	1.5829	41	0.87	0.2338	0.08	[−0.28; 0.44]
MATTIS Conzeptualization	104	0.00	0.0741	41	0.05	0.0337	−0.74	[−1.11; −0.37]
MOCA Abstraction	112	0.53	0.6543	42	1.72	0.5240	−1.90	[−2.31; −1.48]
Random effects model	**320**			**124**			**−0.85**	**[−1.97; 0.27]**
Cohort = PPMI								
MOCA Abstraction	320	−1.90	0.9625	71	1.57	1.5654	−3.16	[−3.50; −2.82]
Random effects model	**1091**			**364**			**−1.04**	**[−2.17; 0.08]**



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Conceptualization across cohorts. The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

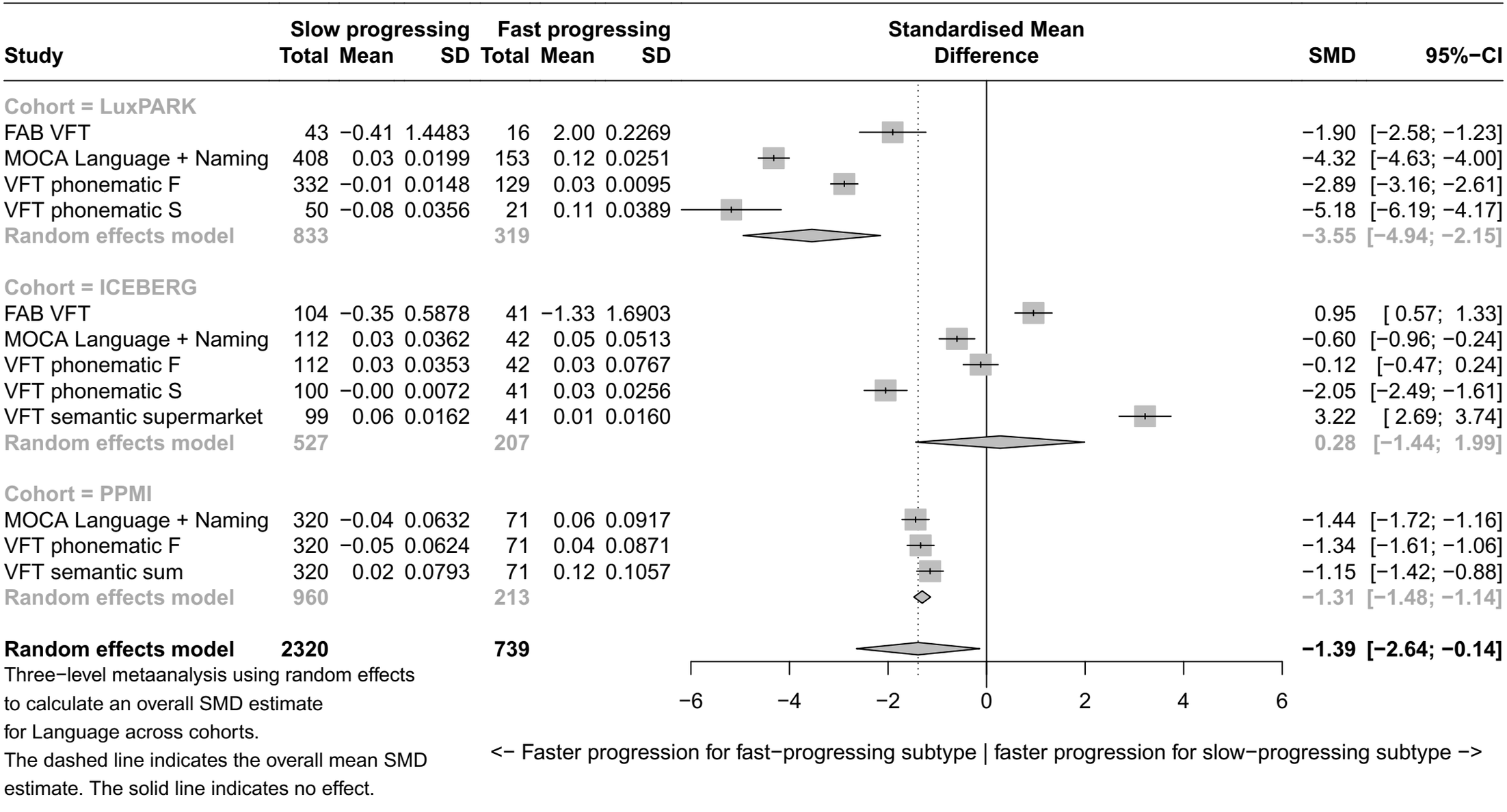
Forest plot for progression characteristics of symptom domain Visuo−executive

Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = ICEBERG								
FAB 3−6	104	0.06	0.0473	41	0.02	0.1003	0.64	[0.27; 1.01]
MATTIS Initiation + Construction	104	0.04	0.0456	41	0.03	0.0611	0.33	[−0.03; 0.69]
MOCA Visuspatial/Executive	112	0.81	0.5368	42	1.37	0.1911	−1.20	[−1.58; −0.82]
Random effects model	320			124			−0.08	[−1.19; 1.04]
Cohort = LuxPARK								
FAB 3−6	42	0.07	0.0238	15	0.09	0.0408	−0.74	[−1.34; −0.13]
MOCA Visuspatial/Executive	408	0.86	0.0973	153	2.12	0.3547	−6.23	[−6.64; −5.82]
Trailmaking A	392	0.02	0.0225	134	0.07	0.0747	−1.17	[−1.38; −0.96]
Trailmaking B	357	0.04	0.0371	99	0.09	0.0523	−1.20	[−1.44; −0.97]
Random effects model	1199			401			−2.34	[−4.89; 0.22]
Cohort = PPMI								
Judgement Line Orientation	320	0.04	0.0357	71	0.17	0.1483	−1.84	[−2.13; −1.55]
MOCA Visuspatial/Executive	320	0.77	1.2703	71	4.01	2.9805	−1.89	[−2.18; −1.60]
Symbol Digit Modalities	320	0.05	0.0727	71	0.21	0.1285	−1.81	[−2.10; −1.52]
Random effects model	960			213			−1.85	[−2.02; −1.68]
Random effects model	**2479**			**738**			**−1.51**	**[−2.67; −0.35]**



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Visuo−executive across cohorts. The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

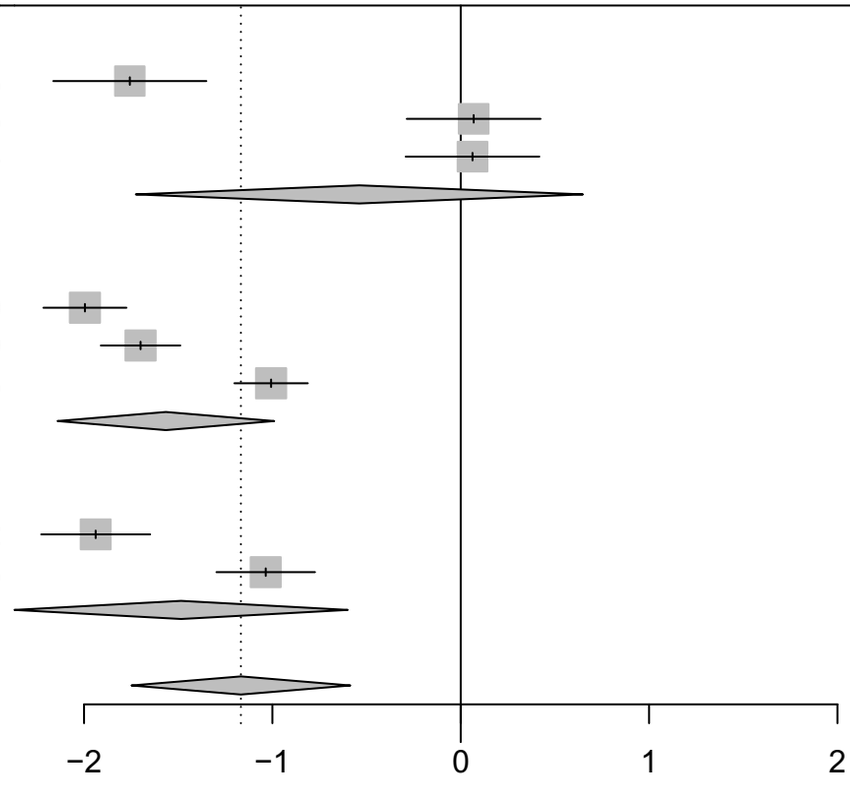
Forest plot for progression characteristics of symptom domain Language



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Language across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

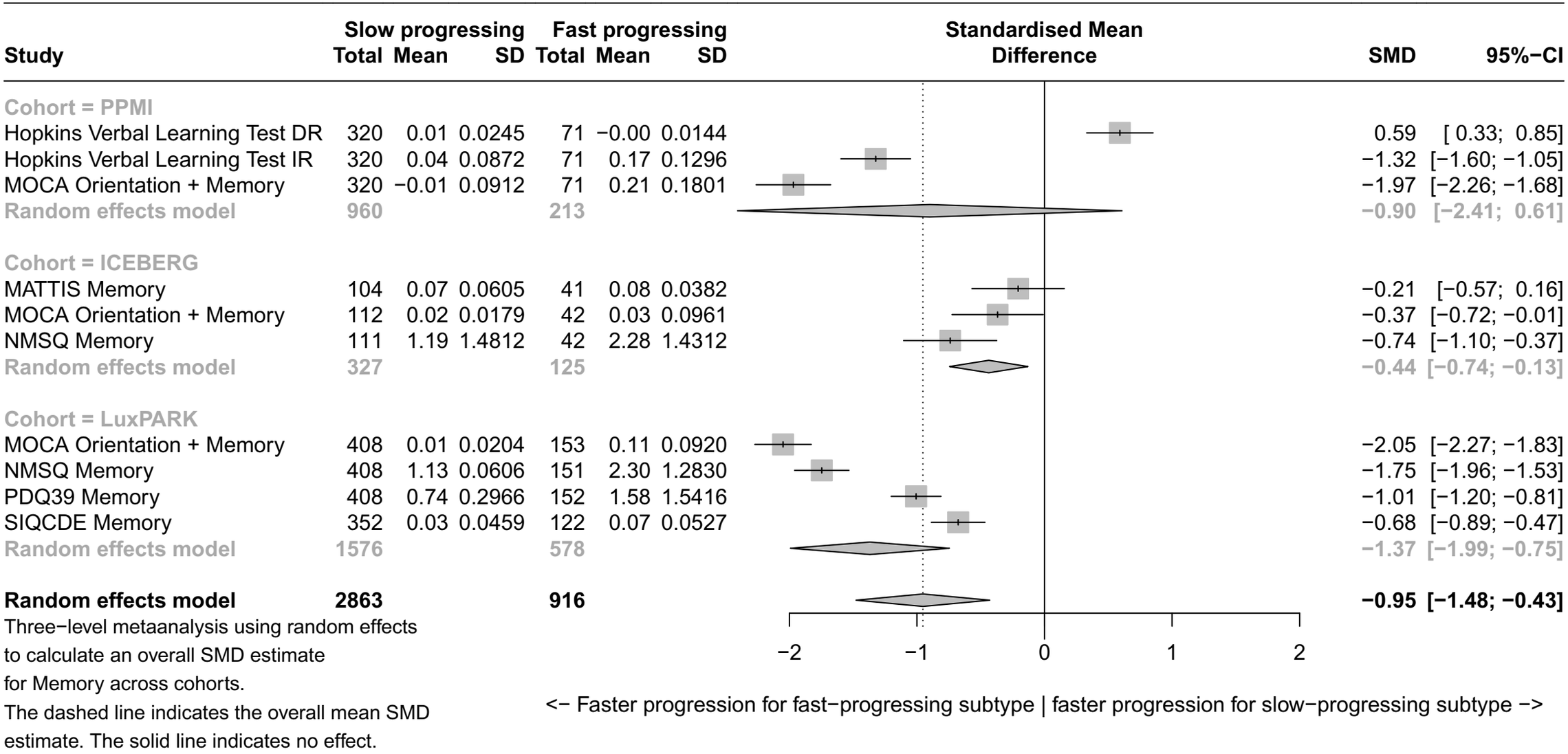
Forest plot for progression characteristics of symptom domain Anxiety

Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = ICEBERG								
HADS anxiety	112	0.02	0.0221	42	0.08	0.0598	−1.76	[−2.16; −1.35]
NMSQ Anxiety	111	1.38	0.8609	42	1.29	1.9033	0.07	[−0.29; 0.42]
UPDRS I Anxiety	112	0.90	0.0762	42	0.84	1.9574	0.06	[−0.29; 0.42]
Random effects model	**335**			**126**			**−0.54**	**[−1.73; 0.65]**
Cohort = LuxPARK								
NMSQ Anxiety	408	1.08	0.3337	151	1.74	0.3269	−2.00	[−2.22; −1.77]
PDQ39 Anxiety	408	1.03	0.3142	153	1.67	0.5120	−1.70	[−1.91; −1.49]
UPDRS I Anxiety	408	0.64	0.1651	153	1.25	1.1298	−1.01	[−1.20; −0.81]
Random effects model	**1224**			**457**			**−1.57**	**[−2.14; −0.99]**
Cohort = PPMI								
STA	323	0.02	0.1127	72	0.23	0.0952	−1.94	[−2.23; −1.65]
UPDRS I Anxiety	335	1.05	2.2489	74	3.37	2.2233	−1.04	[−1.30; −0.77]
Random effects model	**658**			**146**			**−1.48**	**[−2.37; −0.60]**
Random effects model	**2217**			**729**			**−1.17**	**[−1.75; −0.59]**



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Anxiety across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

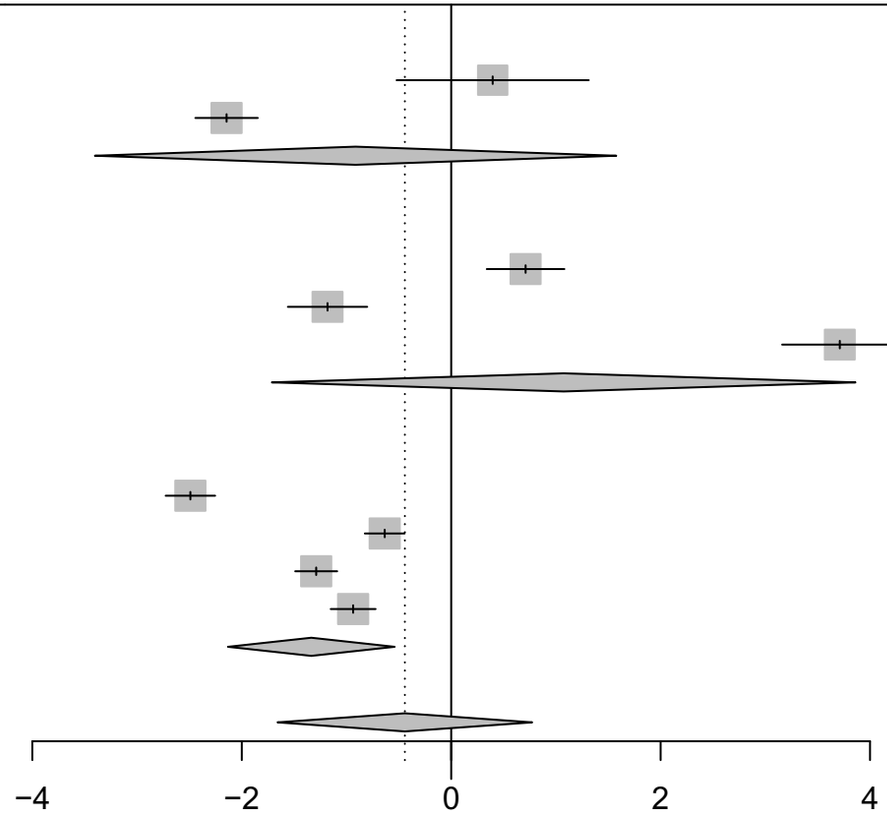
Forest plot for progression characteristics of symptom domain Memory



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Memory across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Attention

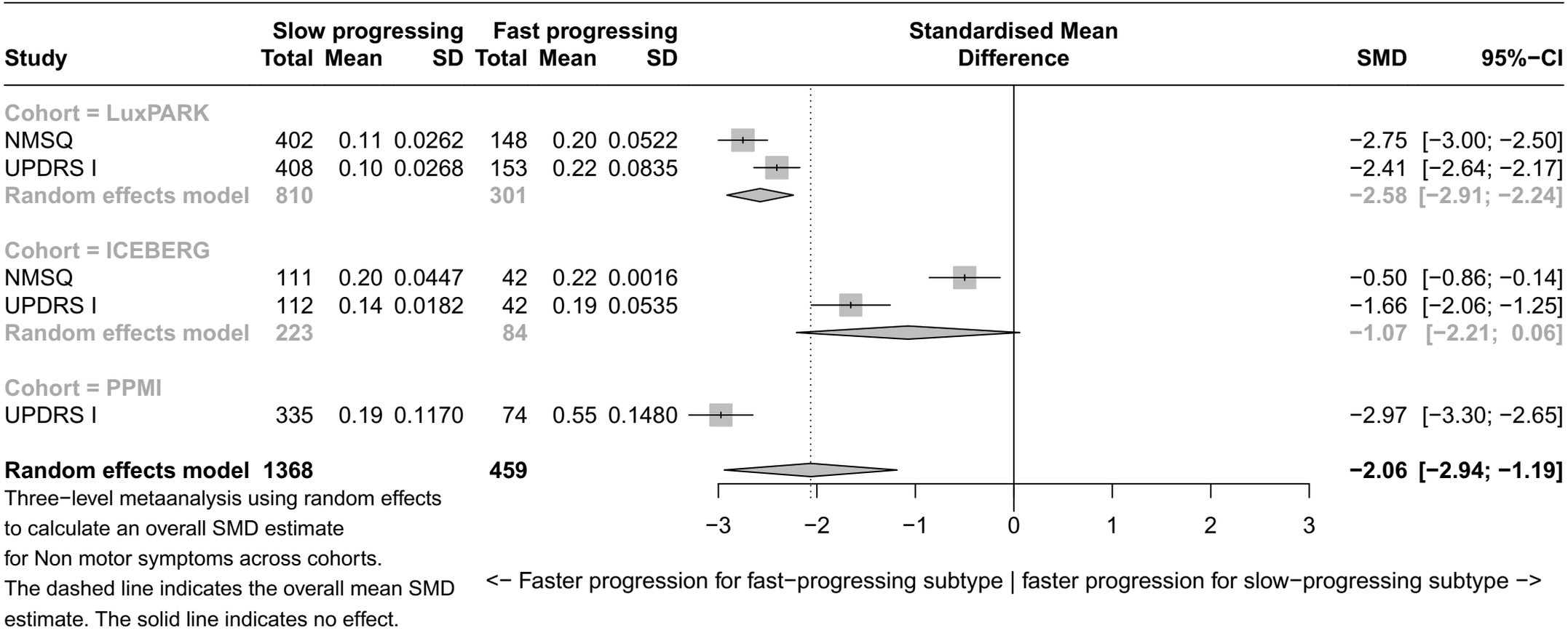
Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = PPMI								
Letter Number Sequencing	56	0.03	0.1650	5	−0.04	0.0202	0.40	[−0.52; 1.31]
MOCA Attention	320	1.33	0.7561	71	3.97	2.4125	−2.15	[−2.44; −1.85]
Random effects model	**376**			**76**			**−0.91**	**[−3.40; 1.58]**
Cohort = ICEBERG								
MATTIS Attention	104	0.00	0.0367	41	−0.02	0.0090	0.71	[0.34; 1.08]
MOCA Attention	112	0.52	0.2299	42	0.94	0.5589	−1.18	[−1.56; −0.80]
NMSQ Attention	111	2.20	0.0771	42	1.79	0.1678	3.71	[3.16; 4.26]
Random effects model	**327**			**125**			**1.07**	**[−1.71; 3.86]**
Cohort = LuxPARK								
MOCA Attention	408	0.58	0.3055	153	1.52	0.5164	−2.49	[−2.73; −2.25]
NMSQ Attention	408	1.47	1.1380	152	2.09	0.1286	−0.63	[−0.82; −0.44]
PDQ39 Attention	408	1.34	0.2652	153	1.77	0.4815	−1.29	[−1.49; −1.09]
SIQCDE Attention	353	0.49	0.3458	122	1.00	0.8884	−0.94	[−1.15; −0.72]
Random effects model	**1577**			**580**			**−1.34**	**[−2.13; −0.54]**
Random effects model	**2280**			**781**			**−0.44**	**[−1.66; 0.77]**



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Attention across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

<− Faster progression for fast−progressing subtype | faster progression for slow−progressing subtype −>

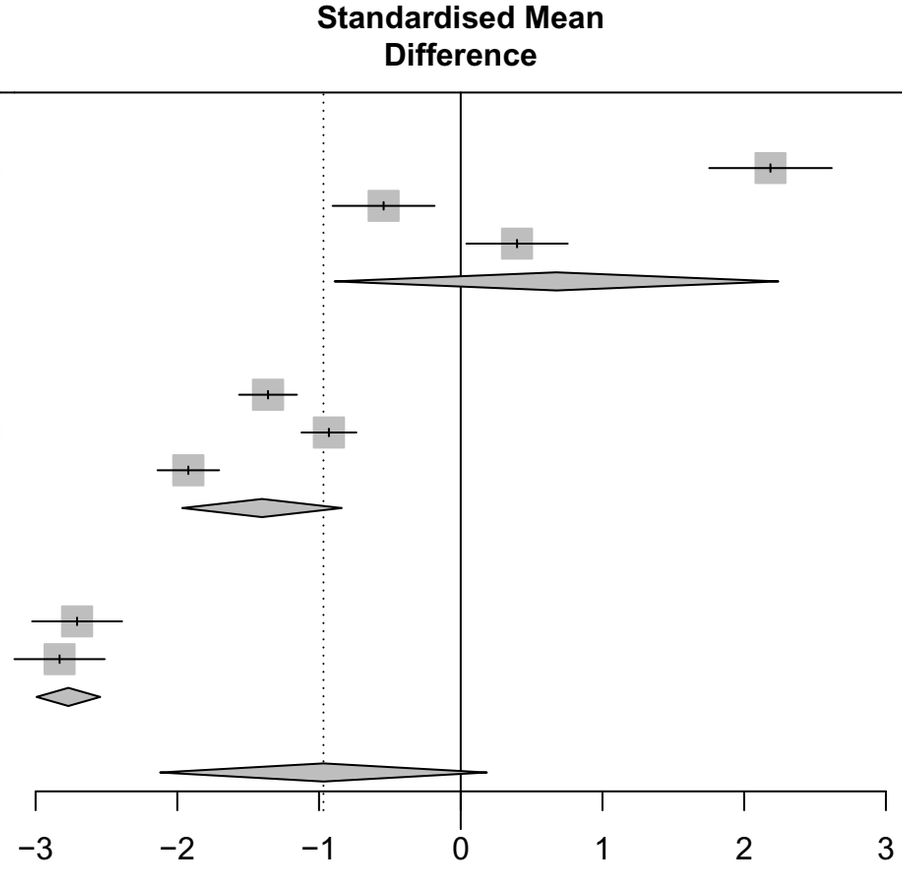
Forest plot for progression characteristics of symptom domain Non motor symptoms



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Non motor symptoms across cohorts. The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

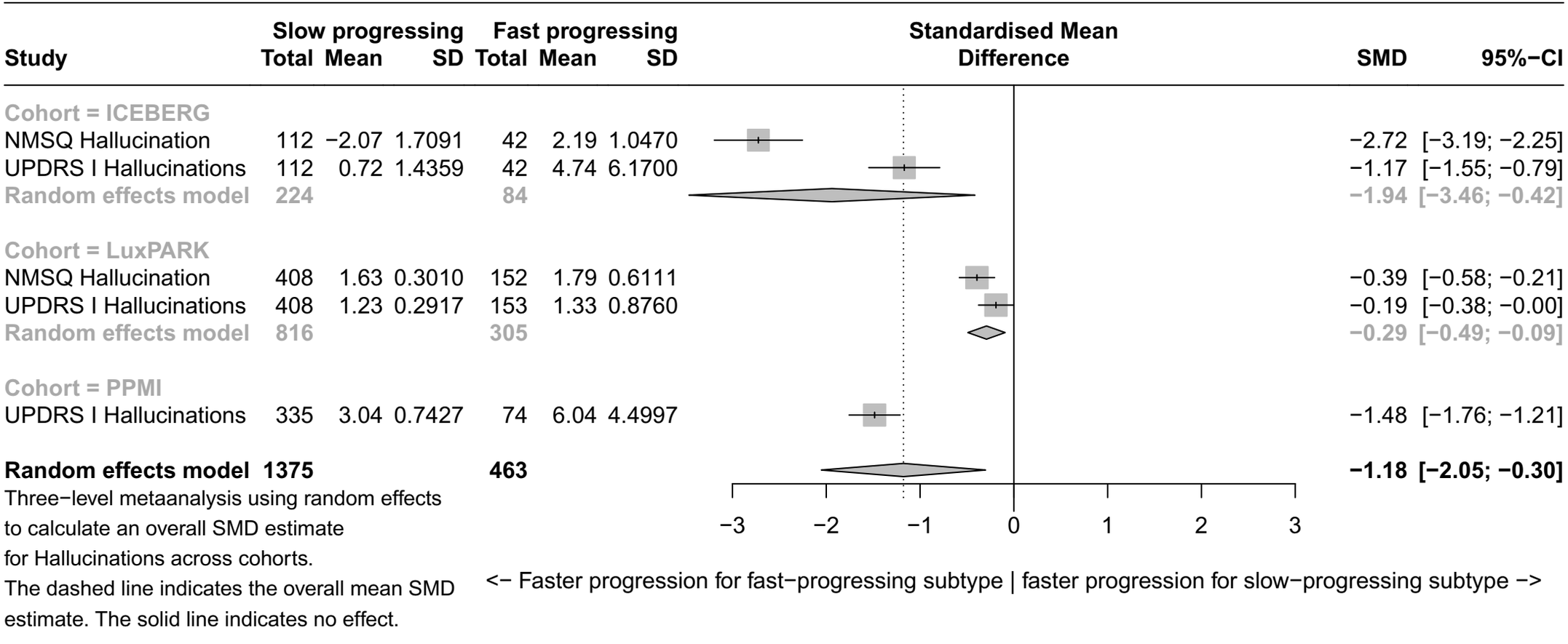
Forest plot for progression characteristics of symptom domain Autonomic

Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = ICEBERG								
NMSQ Autonomic	112	0.20	0.0189	42	0.15	0.0325	2.19	[1.75; 2.62]
SCOPA−AUT	112	0.19	0.0539	42	0.22	0.0462	−0.54	[−0.90; −0.18]
UPDRS I Autonomic	112	0.21	0.0522	42	0.19	0.0746	0.40	[0.04; 0.75]
Random effects model	**336**			**126**			**0.68**	**[−0.89; 2.24]**
Cohort = LuxPARK								
NMSQ Autonomic	403	0.12	0.0237	150	0.17	0.0604	−1.36	[−1.56; −1.16]
SCOPA−AUT	408	0.12	0.0379	153	0.16	0.0429	−0.93	[−1.12; −0.74]
UPDRS I Autonomic	408	0.12	0.0361	153	0.20	0.0612	−1.92	[−2.14; −1.70]
Random effects model	**1219**			**456**			**−1.40**	**[−1.97; −0.84]**
Cohort = PPMI								
SCOPA−AUT	321	0.14	0.0837	73	0.35	0.0108	−2.71	[−3.02; −2.39]
UPDRS I Autonomic	335	0.17	0.1074	74	0.49	0.1282	−2.83	[−3.15; −2.51]
Random effects model	**656**			**147**			**−2.77**	**[−2.99; −2.54]**
Random effects model	**2211**			**729**			**−0.97**	**[−2.12; 0.18]**



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Autonomic across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

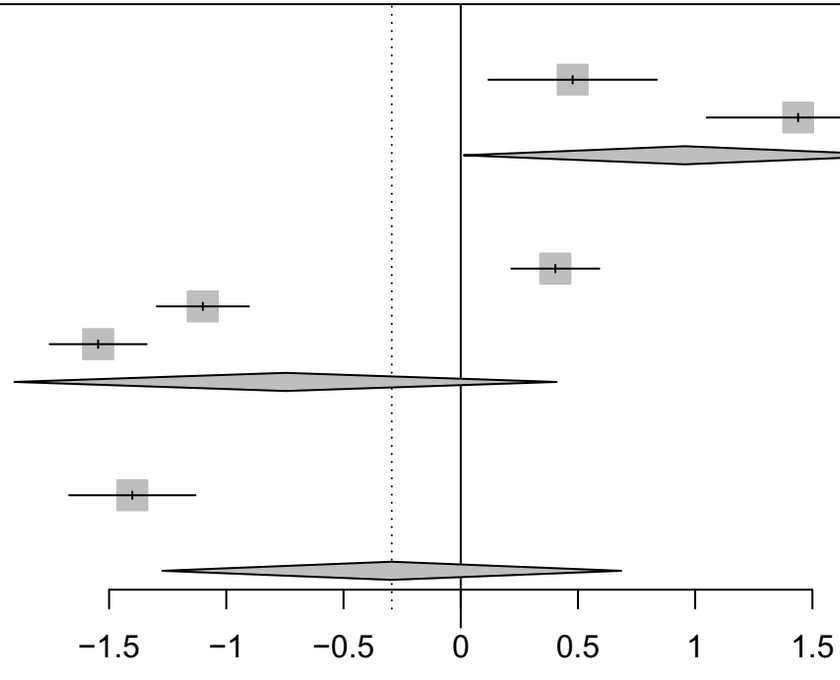
Forest plot for progression characteristics of symptom domain Hallucinations



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Hallucinations across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Pain

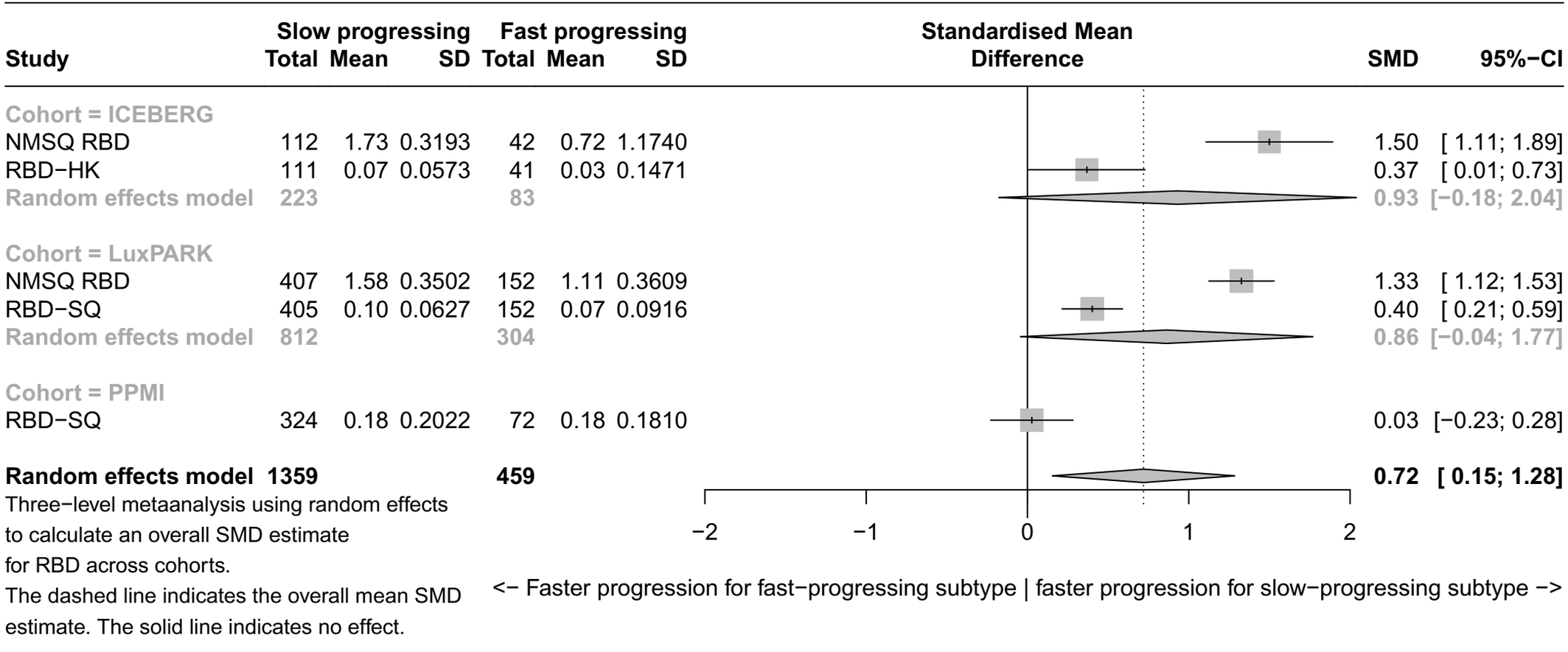
Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = ICEBERG								
NMSQ Pain	111	0.75	0.0044	42	0.59	0.6328	0.48	[0.12; 0.84]
UPDRS I Pain	112	1.04	0.1276	42	0.88	0.0334	1.44	[1.05; 1.83]
Random effects model	**223**			**84**			**0.96**	**[0.01; 1.90]**
Cohort = LuxPARK								
NMSQ Pain	408	0.88	0.1017	152	0.84	0.1111	0.40	[0.22; 0.59]
PDQ39 Pain	407	0.10	0.0208	153	0.14	0.0468	−1.10	[−1.30; −0.90]
UPDRS I Pain	408	1.07	0.1576	153	1.69	0.7281	−1.55	[−1.75; −1.34]
Random effects model	**1223**			**458**			**−0.75**	**[−1.90; 0.41]**
Cohort = PPMI								
UPDRS I Pain	335	1.88	1.6805	74	4.36	2.1147	−1.40	[−1.67; −1.13]
Random effects model	**1781**			**616**			**−0.29**	**[−1.27; 0.68]**



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Pain across cohorts.

The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

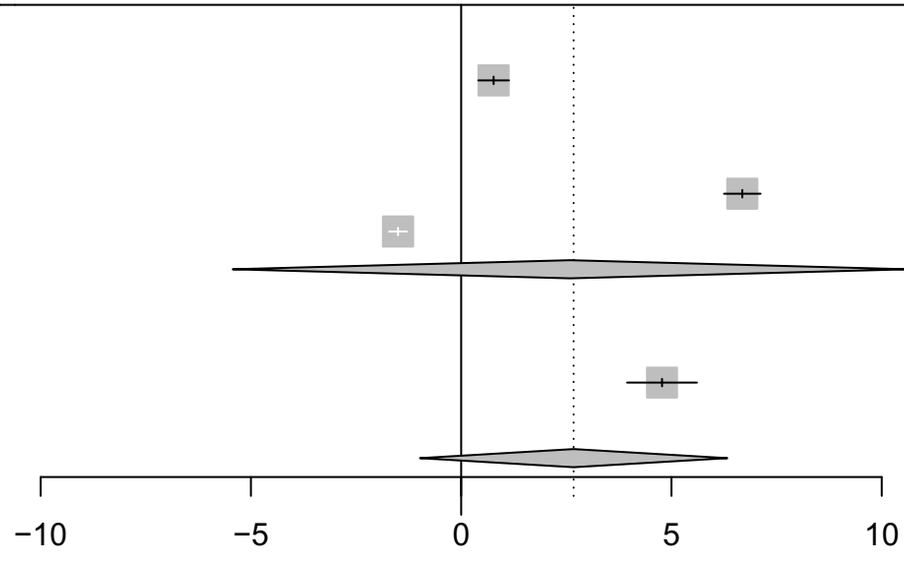
Forest plot for progression characteristics of symptom domain RBD



Three−level metaanalysis using random effects to calculate an overall SMD estimate for RBD across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Smell

Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = ICEBERG								
NMSQ Smell	111	0.38	0.3147	42	−0.33	1.6939	0.77	[0.41; 1.14]
Cohort = LuxPARK								
NMSQ Smell	408	0.62	0.0376	152	0.31	0.0611	6.68	[6.25; 7.12]
Sniffin Test	391	0.04	0.0027	146	0.06	0.0306	−1.50	[−1.71; −1.29]
Random effects model	**799**			**298**			**2.59**	**[−5.43; 10.61]**
Cohort = PPMI								
UPSIT	133	0.19	0.0162	11	0.07	0.0687	4.78	[3.94; 5.61]
Random effects model	**1043**			**351**			**2.67**	**[−0.98; 6.33]**

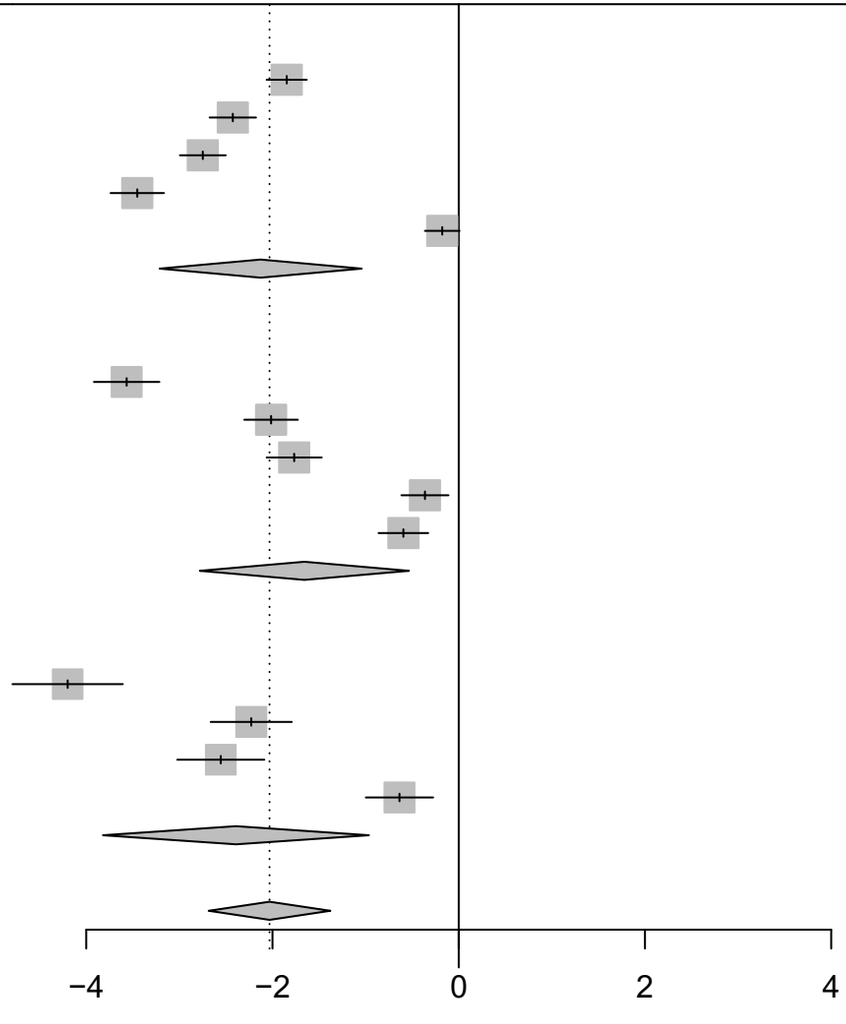


Three−level metaanalysis using random effects to calculate an overall SMD estimate for Smell across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

<− Faster progression for fast−progressing subtype | faster progression for slow−progressing subtype −>

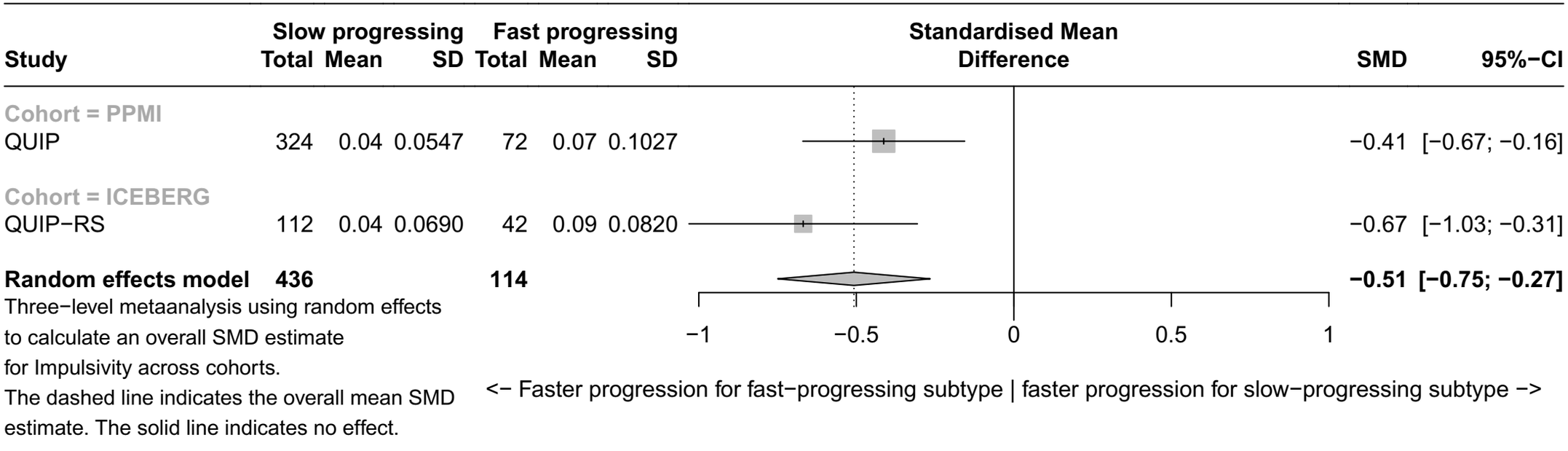
Forest plot for progression characteristics of symptom domain Motor symptoms

Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = LuxPARK								
PDQ39 ADL	408	0.15	0.0652	152	0.36	0.1909	−1.85	[−2.06; −1.63]
Pegboard	386	0.06	0.0171	128	0.12	0.0374	−2.43	[−2.68; −2.18]
UPDRS II	408	0.15	0.0555	153	0.42	0.1606	−2.75	[−2.99; −2.50]
UPDRS III on	380	0.09	0.0212	139	0.23	0.0707	−3.45	[−3.74; −3.17]
UPDRS IV	408	0.09	0.0759	153	0.11	0.0872	−0.18	[−0.36; 0.01]
Random effects model	**1990**			**725**			**−2.13**	**[−3.21; −1.04]**
Cohort = PPMI								
UPDRS II	335	0.17	0.1047	74	0.59	0.1604	−3.57	[−3.92; −3.21]
UPDRS III off	335	0.26	0.1491	74	0.58	0.1757	−2.02	[−2.30; −1.73]
UPDRS III on	300	0.03	0.1291	66	0.31	0.2480	−1.77	[−2.06; −1.47]
UPDRS III untreated	335	0.55	0.1221	74	0.61	0.3608	−0.36	[−0.62; −0.11]
UPDRS IV	301	0.26	0.1800	68	0.37	0.2256	−0.59	[−0.86; −0.33]
Random effects model	**1606**			**356**			**−1.66**	**[−2.78; −0.53]**
Cohort = ICEBERG								
UPDRS II	112	0.20	0.0313	42	0.39	0.0694	−4.20	[−4.79; −3.61]
UPDRS III off	111	0.04	0.0082	42	0.16	0.0983	−2.23	[−2.66; −1.79]
UPDRS III on	105	0.01	0.0023	41	0.09	0.0623	−2.55	[−3.02; −2.09]
UPDRS IV	112	0.07	0.1174	42	0.16	0.1752	−0.64	[−1.00; −0.27]
Random effects model	**440**			**167**			**−2.39**	**[−3.82; −0.96]**
Random effects model	**4036**			**1248**			**−2.03**	**[−2.69; −1.38]**



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Motor symptoms across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Impulsivity

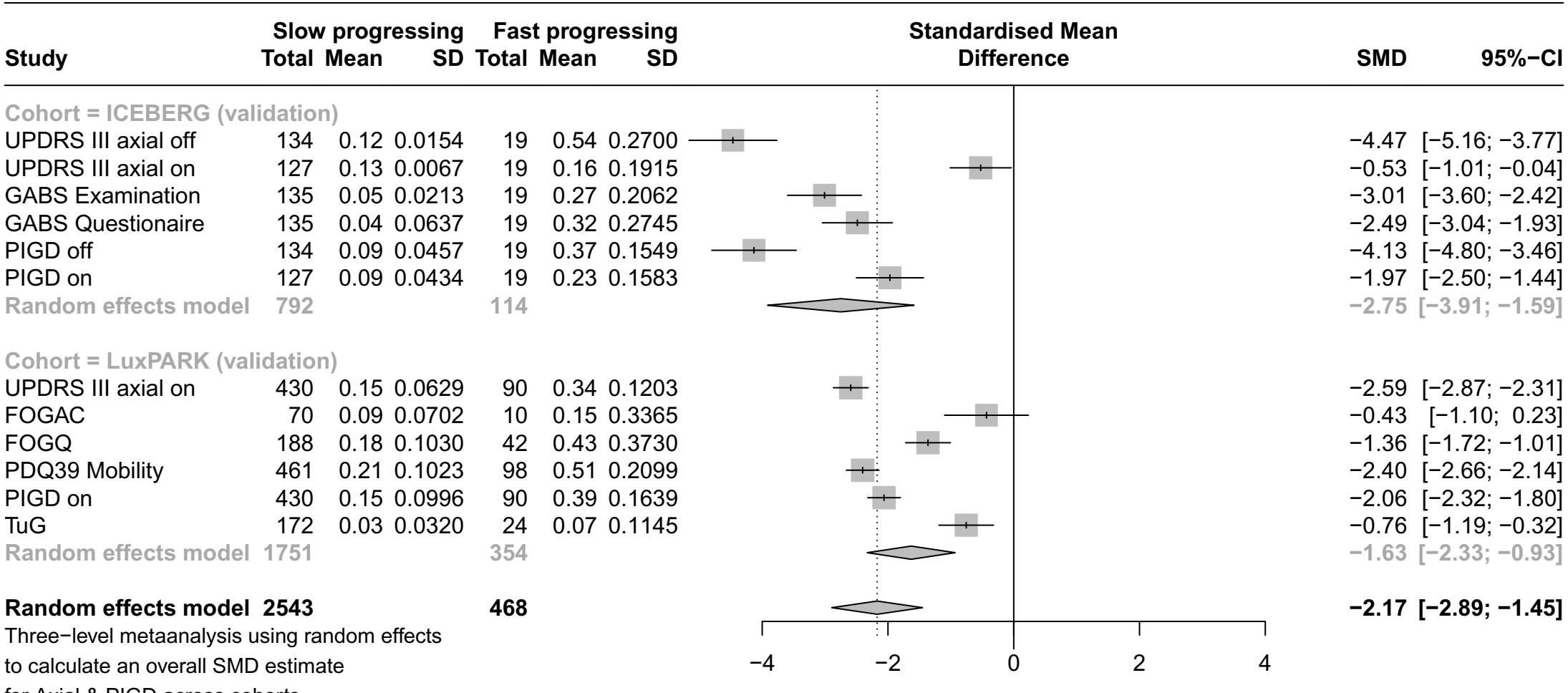


Three−level metaanalysis using random effects to calculate an overall SMD estimate for Impulsivity across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Progression subtypes in Parkinson’s disease identified by a data driven multi cohort analysis

Forest plots for symptom domain progression (cross-cohort validation)

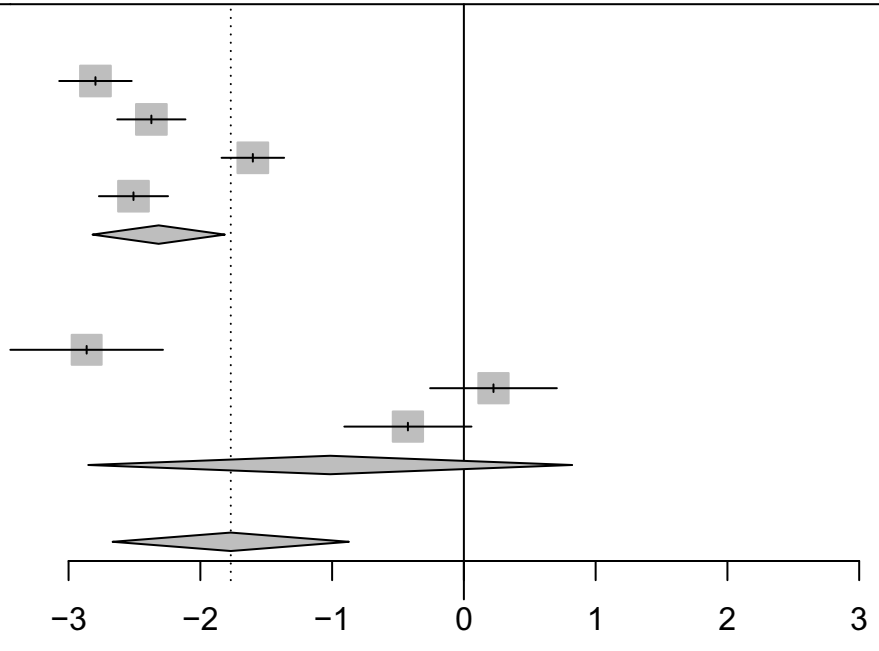
Forest plot for progression characteristics of symptom domain Axial & PIGD (validation)



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Axial & PIGD across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

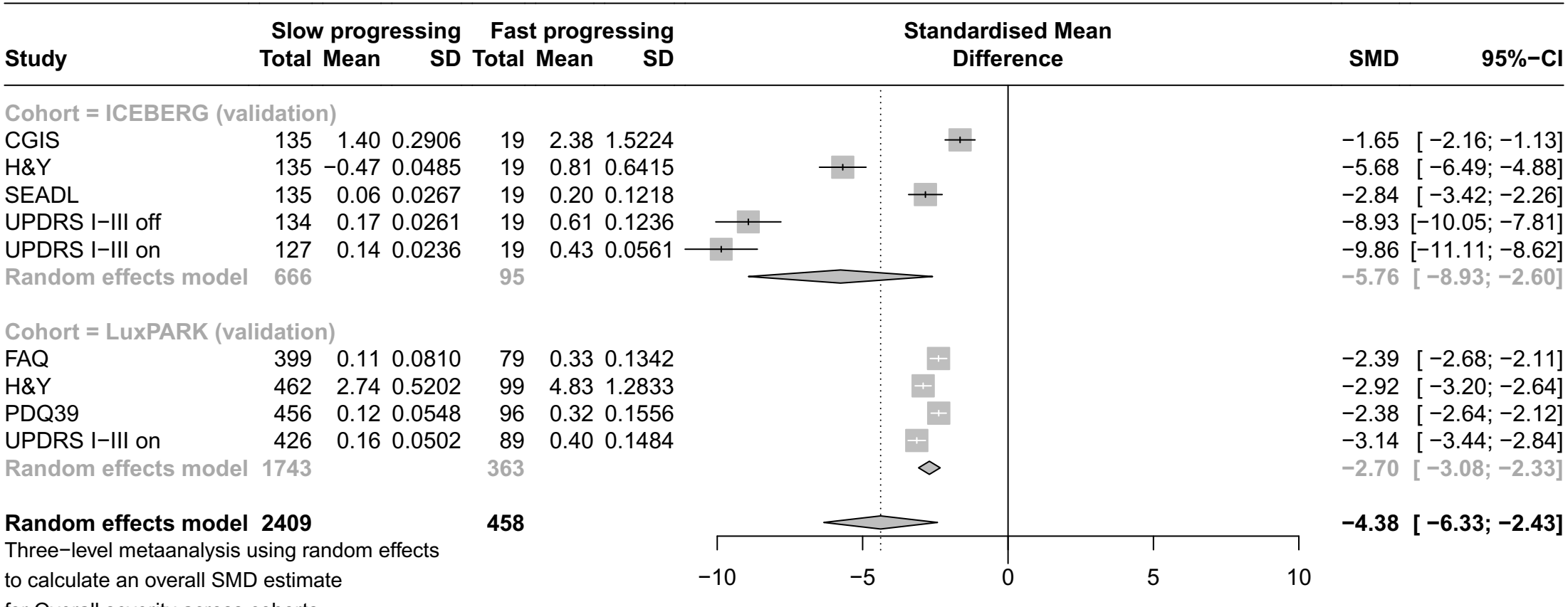
Forest plot for progression characteristics of symptom domain Depression (validation)

Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = LuxPARK (validation)								
BDI	459	0.06	0.0305	96	0.14	0.0120	−2.80	[−3.07; −2.52]
NMSQ Depression	462	1.19	0.4669	98	2.22	0.2138	−2.37	[−2.63; −2.11]
PDQ39 Depression	461	0.08	0.0500	98	0.16	0.0658	−1.60	[−1.84; −1.36]
UPDRS I Depression	462	0.92	0.2209	99	1.83	0.7230	−2.51	[−2.77; −2.25]
Random effects model	**1844**			**391**			**−2.32**	**[−2.82; −1.81]**
Cohort = ICEBERG (validation)								
HADS depression	135	0.11	0.0026	19	0.17	0.0615	−2.86	[−3.44; −2.28]
NMSQ Depression	135	1.83	0.2900	19	1.59	2.8964	0.22	[−0.26; 0.71]
UPDRS I Depression	135	1.32	0.7801	19	1.64	0.6916	−0.42	[−0.91; 0.06]
Random effects model	**405**			**57**			**−1.01**	**[−2.85; 0.82]**
Random effects model	**2249**			**448**			**−1.77**	**[−2.67; −0.87]**



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Depression across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

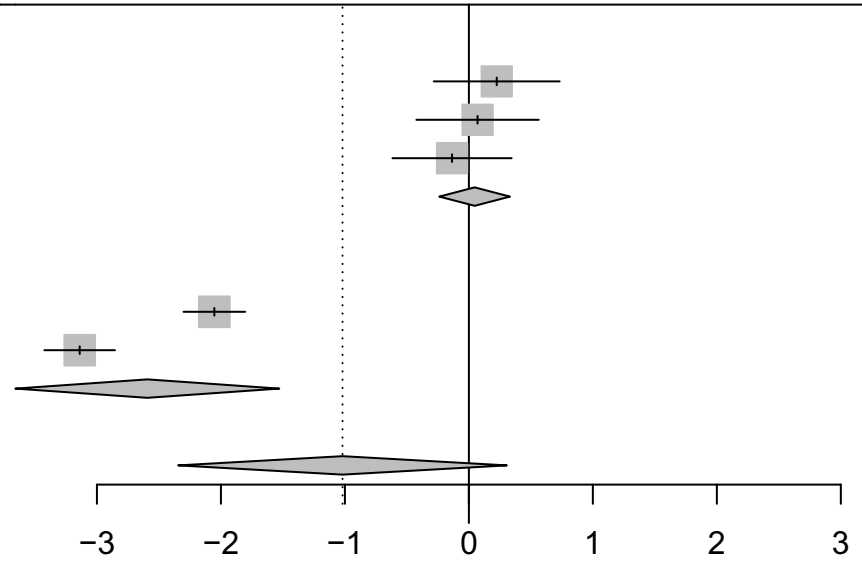
Forest plot for progression characteristics of symptom domain Overall severity (validation)



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Overall severity across cohorts. The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Apathy (validation)

Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = ICEBERG (validation)								
DAS	124	0.09	0.0128	17	0.09	0.0490	0.22	[−0.28; 0.73]
SAS	127	0.09	0.0224	18	0.09	0.0462	0.07	[−0.42; 0.56]
UPDRS I Apathy	135	1.11	0.1876	19	1.17	1.1351	−0.14	[−0.62; 0.34]
Random effects model	**386**			**54**			**0.05**	**[−0.24; 0.33]**
Cohort = LuxPARK (validation)								
SAS	459	0.05	0.0433	99	0.14	0.0457	−2.05	[−2.30; −1.80]
UPDRS I Apathy	462	0.88	0.0486	99	2.08	0.9082	−3.14	[−3.42; −2.85]
Random effects model	**921**			**198**			**−2.59**	**[−3.66; −1.53]**
Random effects model	**1307**			**252**			**−1.02**	**[−2.34; 0.31]**

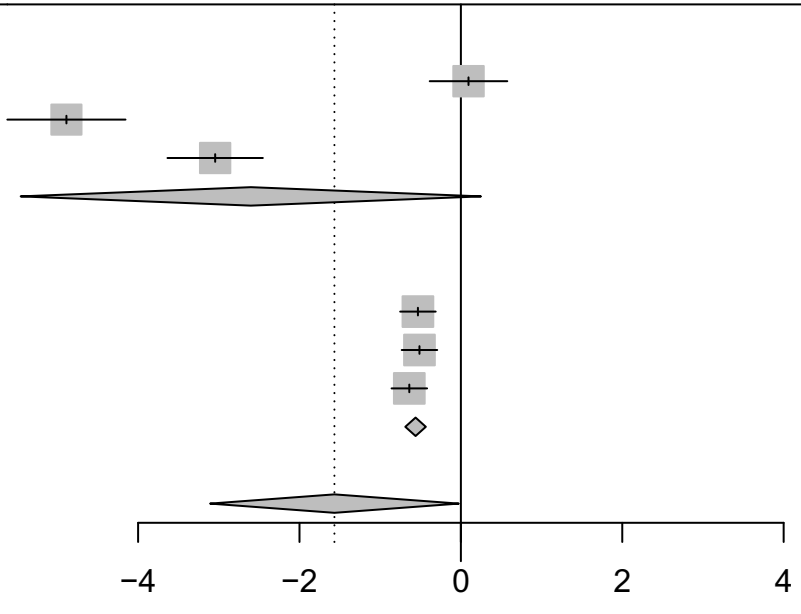


Three−level metaanalysis using random effects to calculate an overall SMD estimate for Apathy across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

<− Faster progression for fast−progressing subtype | faster progression for slow−progressing subtype −>

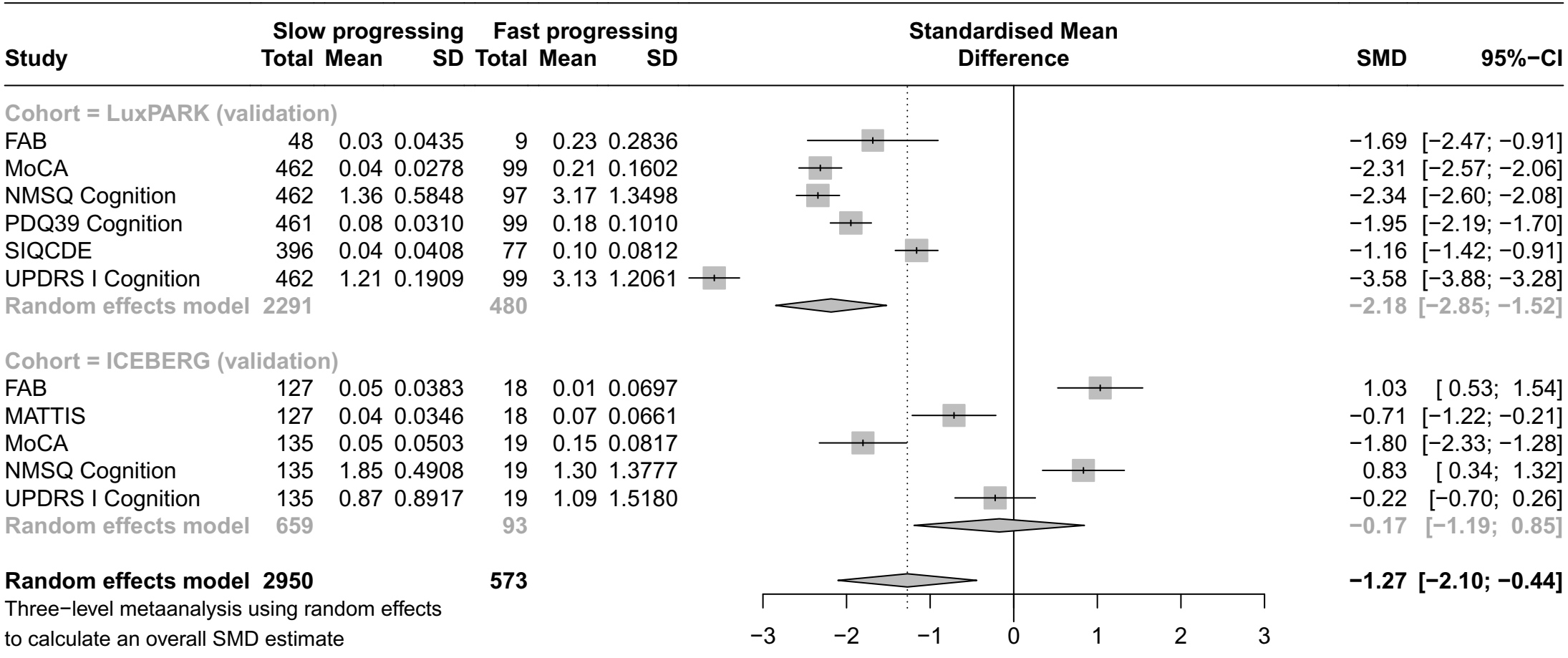
Forest plot for progression characteristics of symptom domain Sleep (validation)

Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = ICEBERG (validation)								
ESS	133	0.17	0.1063	19	0.16	0.0548	0.09	[−0.39; 0.58]
NMSQ Sleep	135	1.21	0.2211	19	3.35	1.1076	−4.89	[−5.62; −4.16]
UPDRS I Sleep	135	0.08	0.0224	19	0.15	0.0255	−3.05	[−3.64; −2.45]
Random effects model	**403**			**57**			**−2.60**	**[−5.46; 0.25]**
Cohort = LuxPARK (validation)								
NMSQ Sleep	462	0.68	0.0575	98	0.76	0.3560	−0.53	[−0.75; −0.31]
PDSS	460	0.11	0.0237	98	0.12	0.0355	−0.51	[−0.73; −0.29]
UPDRS I Sleep	462	0.10	0.0325	99	0.12	0.0353	−0.64	[−0.86; −0.42]
Random effects model	**1384**			**295**			**−0.56**	**[−0.69; −0.43]**
Random effects model	**1787**			**352**			**−1.57**	**[−3.11; −0.03]**



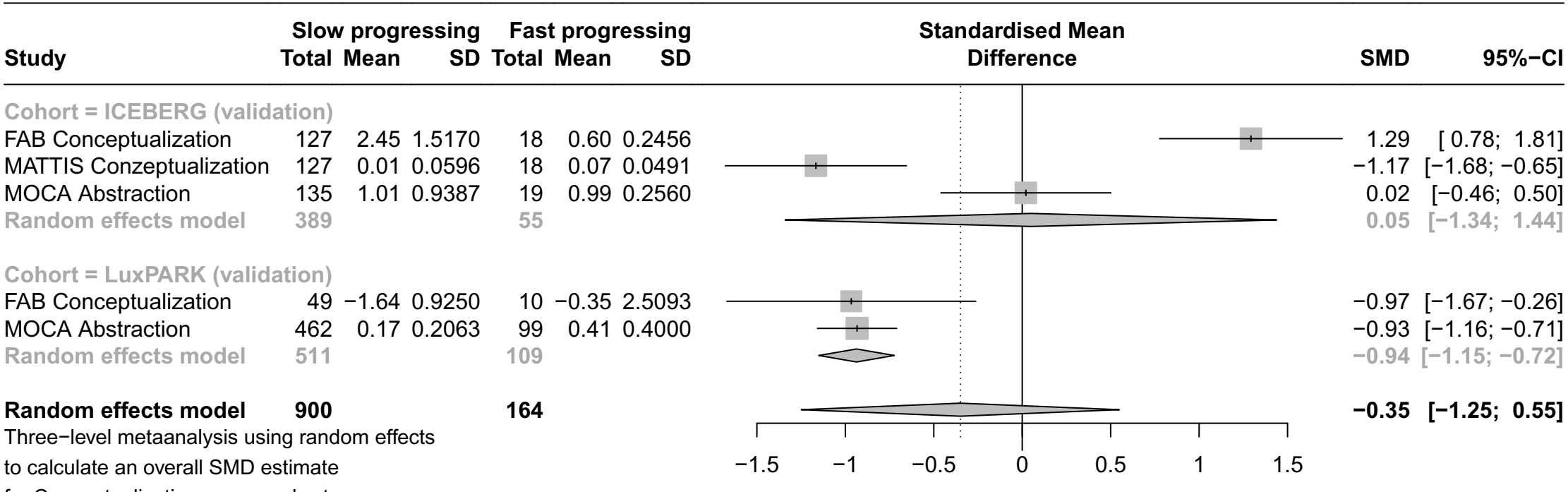
Three−level metaanalysis using random effects to calculate an overall SMD estimate for Sleep across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Overall cognition (validation)



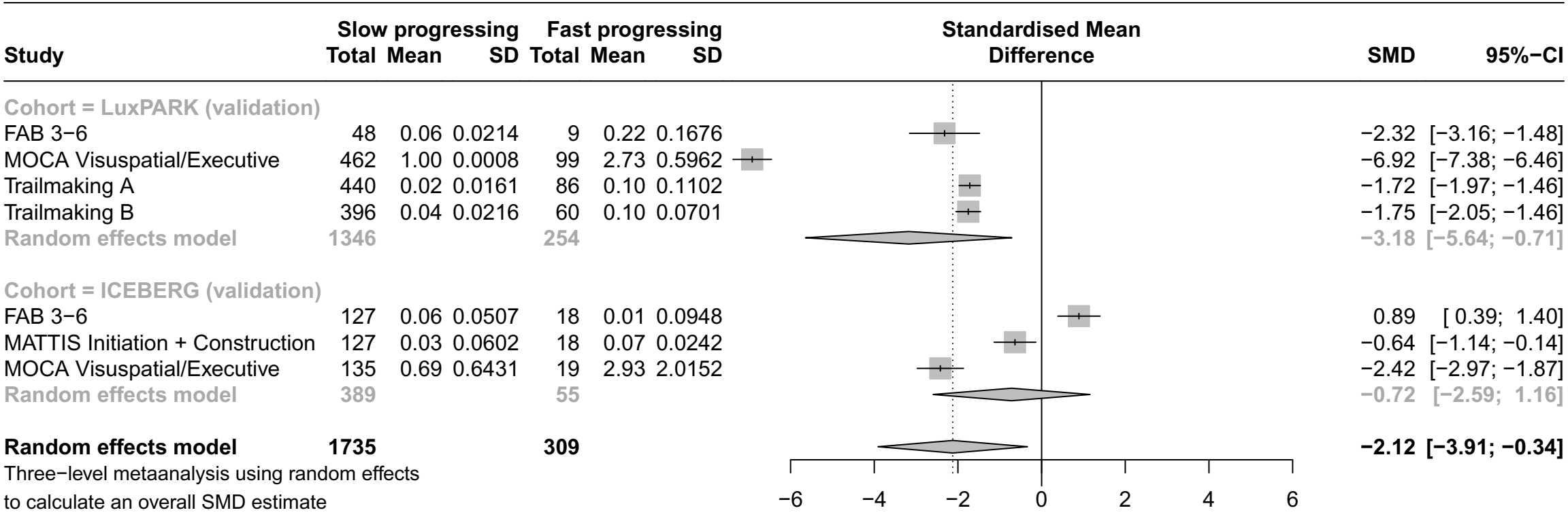
Three−level metaanalysis using random effects to calculate an overall SMD estimate for Overall cognition across cohorts. The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Conceptualization (validation)



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Conceptualization across cohorts. The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

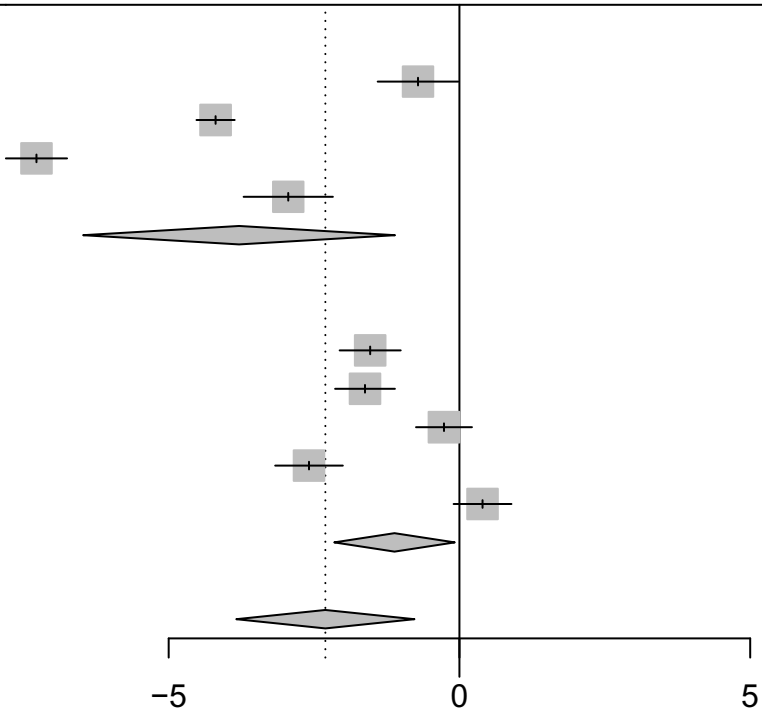
Forest plot for progression characteristics of symptom domain Visuo−executive (validation)



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Visuo−executive across cohorts. The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

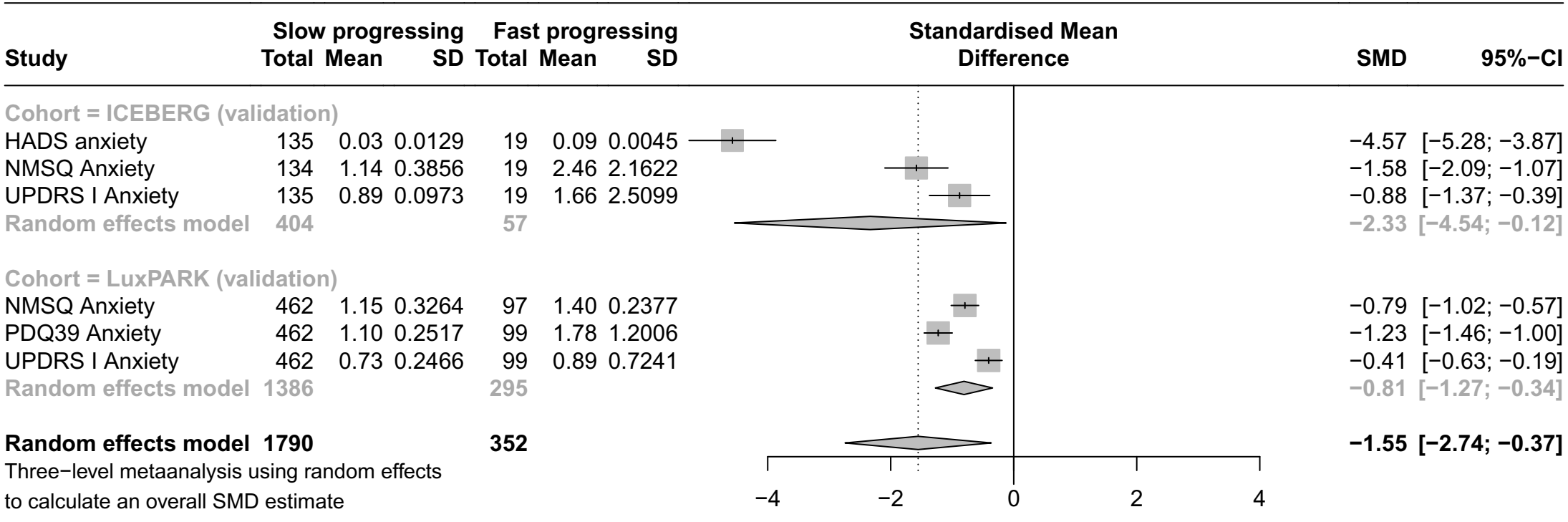
Forest plot for progression characteristics of symptom domain Language (validation)

Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = LuxPARK (validation)								
FAB VFT	49	0.61	2.3727	10	2.27	1.7866	−0.71	[−1.41; −0.02]
MOCA Language + Naming	462	0.04	0.0173	99	0.18	0.0729	−4.19	[−4.52; −3.86]
VFT phonematic F	374	−0.00	0.0089	87	0.06	0.0011	−7.27	[−7.80; −6.75]
VFT phonematic S	57	−0.02	0.0422	14	0.17	0.1257	−2.94	[−3.71; −2.18]
Random effects model	**942**			**210**			**−3.79**	**[−6.47; −1.11]**
Cohort = ICEBERG (validation)								
FAB VFT	127	−0.43	0.5594	18	0.38	0.0389	−1.53	[−2.06; −1.01]
MOCA Language + Naming	135	0.03	0.0221	19	0.08	0.0620	−1.62	[−2.14; −1.11]
VFT phonematic F	135	0.03	0.0352	19	0.04	0.0665	−0.27	[−0.75; 0.22]
VFT phonematic S	123	0.00	0.0105	18	0.03	0.0003	−2.59	[−3.17; −2.01]
VFT semantic supermarket	122	0.05	0.0317	18	0.04	0.0017	0.40	[−0.10; 0.89]
Random effects model	**642**			**92**			**−1.12**	**[−2.15; −0.08]**
Random effects model	**1584**			**302**			**−2.31**	**[−3.84; −0.77]**



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Language across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

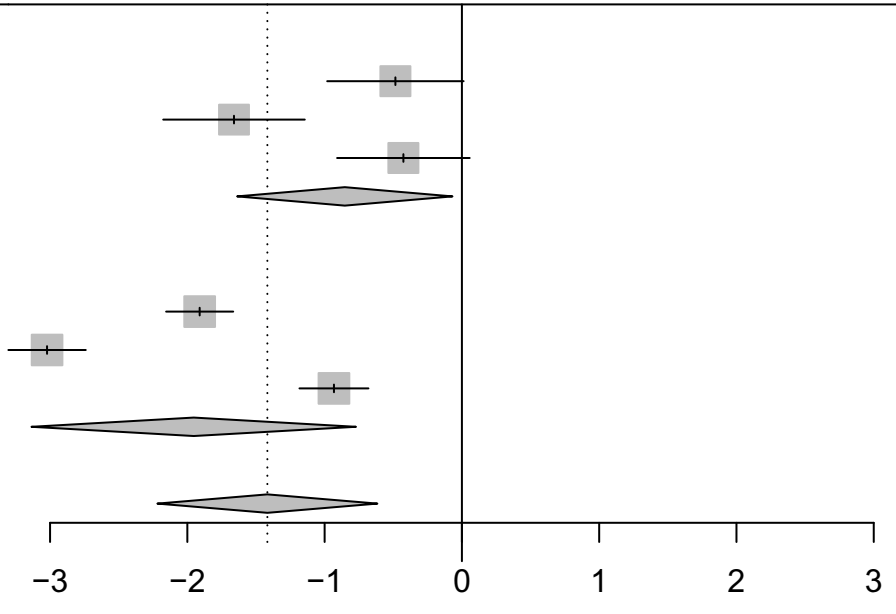
Forest plot for progression characteristics of symptom domain Anxiety (validation)



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Anxiety across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Memory (validation)

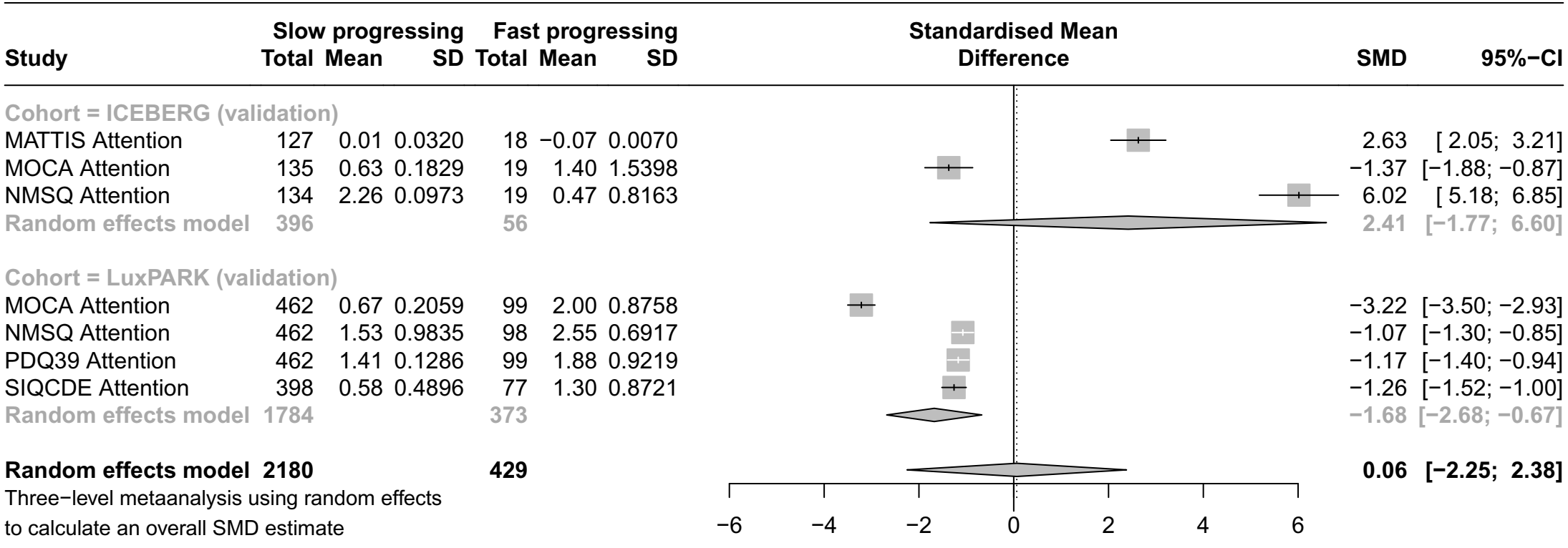
Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = ICEBERG (validation)								
MATTIS Memory	127	0.07	0.0498	18	0.09	0.0571	−0.48	[−0.98; 0.01]
MOCA Orientation + Memory	135	0.01	0.0232	19	0.06	0.0612	−1.66	[−2.18; −1.14]
NMSQ Memory	134	1.37	1.2394	19	1.95	2.0623	−0.43	[−0.91; 0.06]
Random effects model	**396**			**56**			**−0.85**	**[−1.64; −0.07]**
Cohort = LuxPARK (validation)								
MOCA Orientation + Memory	462	0.01	0.0199	99	0.14	0.1483	−1.91	[−2.15; −1.67]
NMSQ Memory	462	1.18	0.1844	97	2.77	1.1996	−3.02	[−3.30; −2.74]
SIQCDE Memory	397	0.04	0.0417	77	0.08	0.0691	−0.93	[−1.18; −0.68]
Random effects model	**1321**			**273**			**−1.95**	**[−3.14; −0.77]**
Random effects model	**1717**			**329**			**−1.42**	**[−2.22; −0.62]**



<− Faster progression for fast−progressing subtype | faster progression for slow−progressing subtype −>

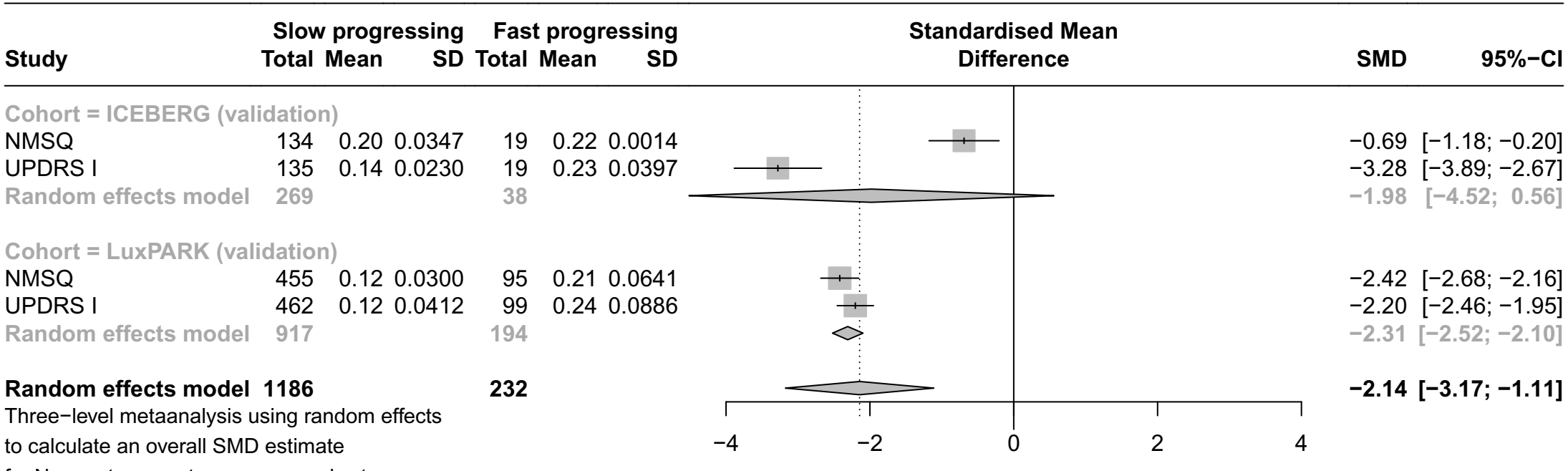
Three−level metaanalysis using random effects to calculate an overall SMD estimate for Memory across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Attention (validation)



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Attention across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

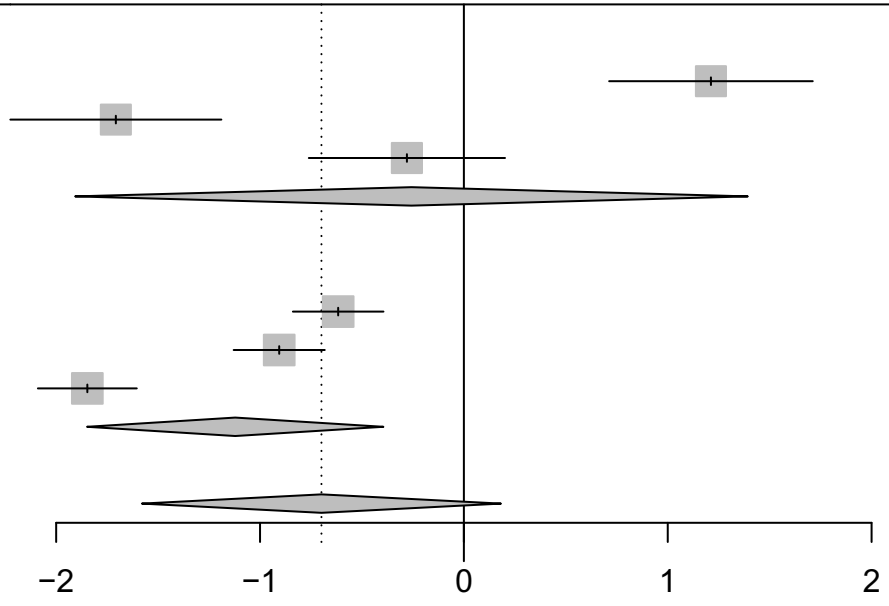
Forest plot for progression characteristics of symptom domain Non motor symptoms (validation)



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Non motor symptoms across cohorts. The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Autonomic (validation)

Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = ICEBERG (validation)								
NMSQ Autonomic	135	0.19	0.0132	19	0.16	0.0734	1.21	[0.71; 1.71]
SCOPA−AUT	135	0.19	0.0481	19	0.28	0.0830	−1.71	[−2.22; −1.19]
UPDRS I Autonomic	135	0.20	0.0600	19	0.21	0.0328	−0.28	[−0.76; 0.20]
Random effects model	**405**			**57**			**−0.26**	**[−1.91; 1.39]**
Cohort = LuxPARK (validation)								
NMSQ Autonomic	456	0.13	0.0288	97	0.15	0.0716	−0.62	[−0.84; −0.39]
SCOPA−AUT	462	0.13	0.0363	99	0.17	0.0566	−0.91	[−1.13; −0.68]
UPDRS I Autonomic	462	0.13	0.0428	99	0.22	0.0693	−1.85	[−2.09; −1.60]
Random effects model	**1380**			**295**			**−1.12**	**[−1.85; −0.39]**
Random effects model	**1785**			**352**			**−0.70**	**[−1.58; 0.18]**

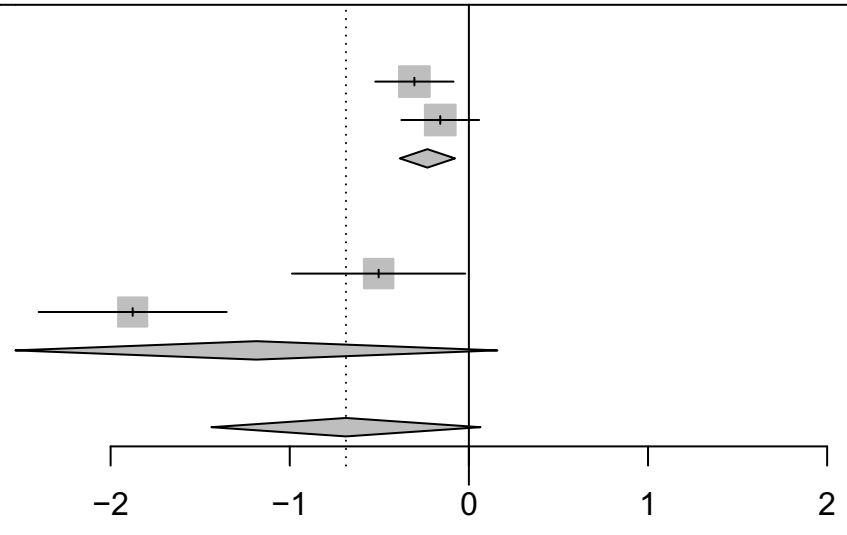


Three−level metaanalysis using random effects to calculate an overall SMD estimate for Autonomic across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

<− Faster progression for fast−progressing subtype | faster progression for slow−progressing subtype −>

Forest plot for progression characteristics of symptom domain Hallucinations (validation)

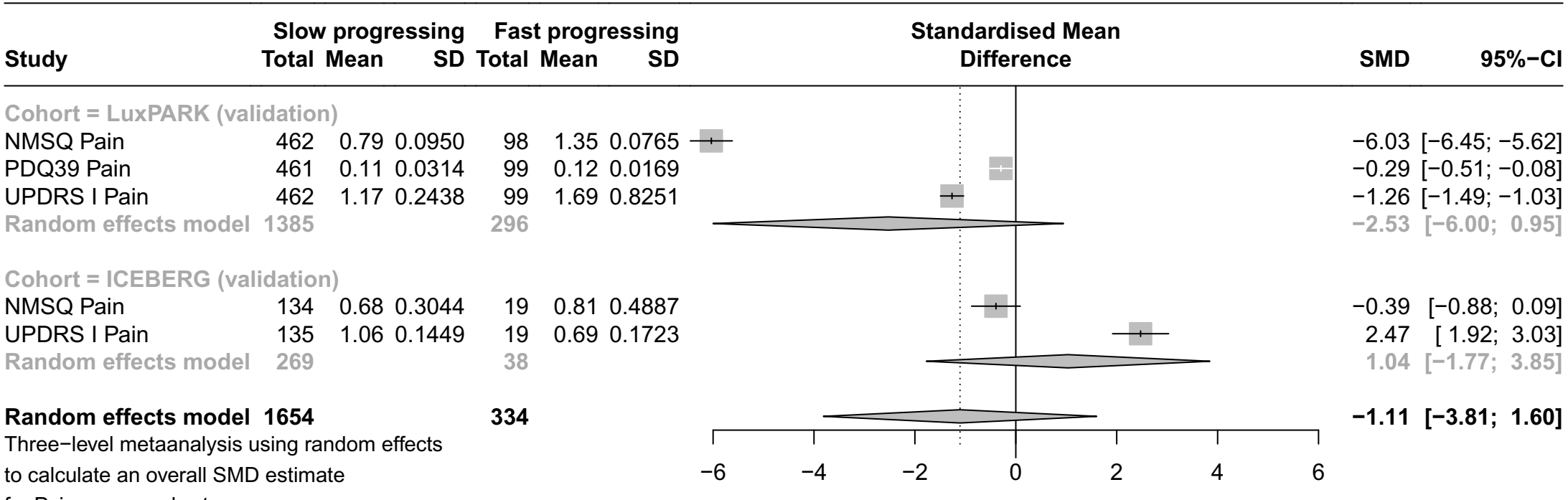
Study	Slow progressing Total	Mean	SD	Fast progressing Total	Mean	SD	SMD	95%−CI
Cohort = LuxPARK (validation)								
NMSQ Hallucination	462	1.67	0.2022	98	1.81	1.0142	−0.30	[−0.52; −0.09]
UPDRS I Hallucinations	462	1.31	0.2786	99	1.38	0.8747	−0.16	[−0.38; 0.06]
Random effects model	**924**			**197**			**−0.23**	**[−0.39; −0.08]**
Cohort = ICEBERG (validation)								
NMSQ Hallucination	135	0.93	0.9204	19	1.38	0.7406	−0.50	[−0.99; −0.02]
UPDRS I Hallucinations	135	1.03	2.0070	19	8.38	9.9115	−1.88	[−2.40; −1.35]
Random effects model	**270**			**38**			**−1.19**	**[−2.53; 0.16]**
Random effects model	**1194**			**235**			**−0.69**	**[−1.44; 0.06]**



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Hallucinations across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

<− Faster progression for fast−progressing subtype | faster progression for slow−progressing subtype −>

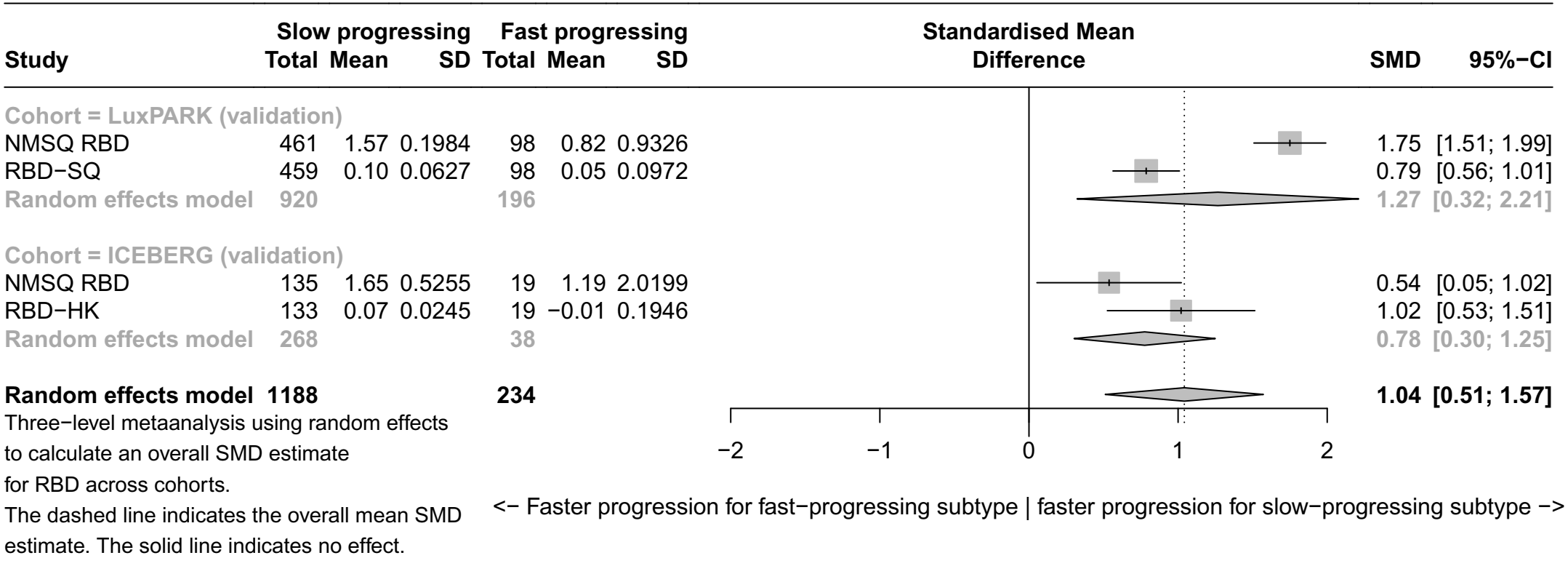
Forest plot for progression characteristics of symptom domain Pain (validation)



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Pain across cohorts.

The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

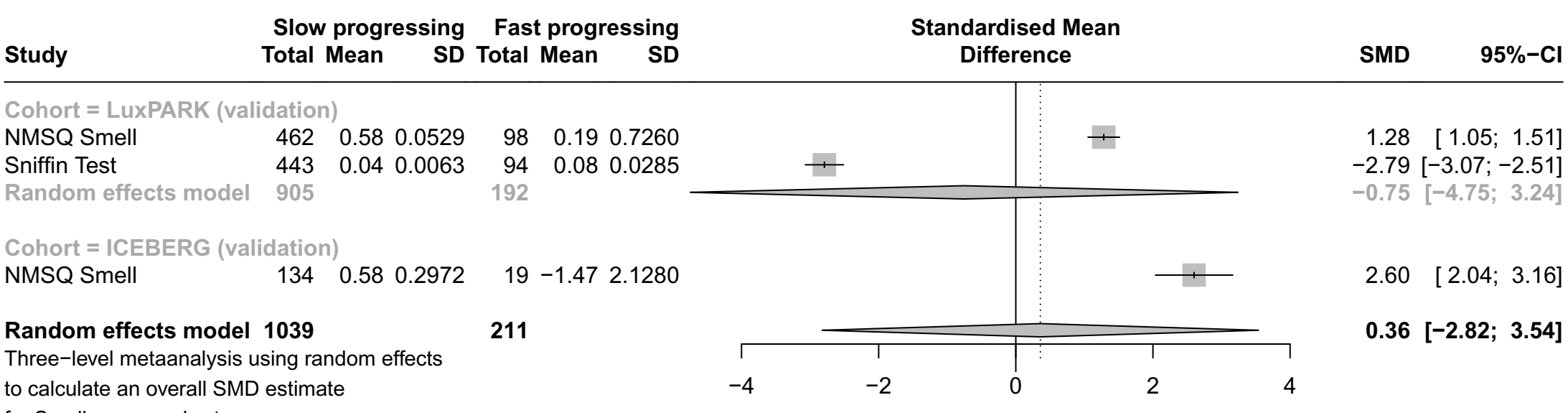
Forest plot for progression characteristics of symptom domain RBD (validation)



Three−level metaanalysis using random effects to calculate an overall SMD estimate for RBD across cohorts.

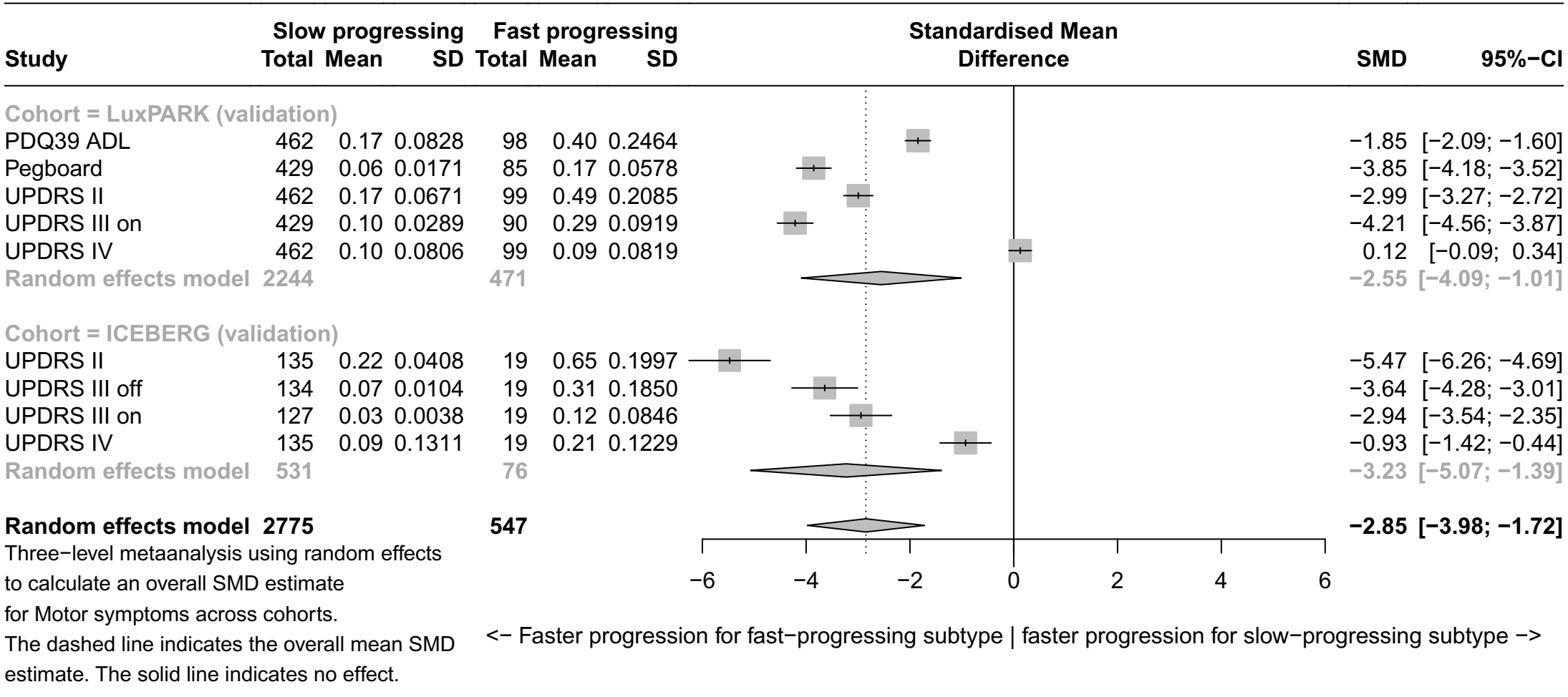
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Smell (validation)



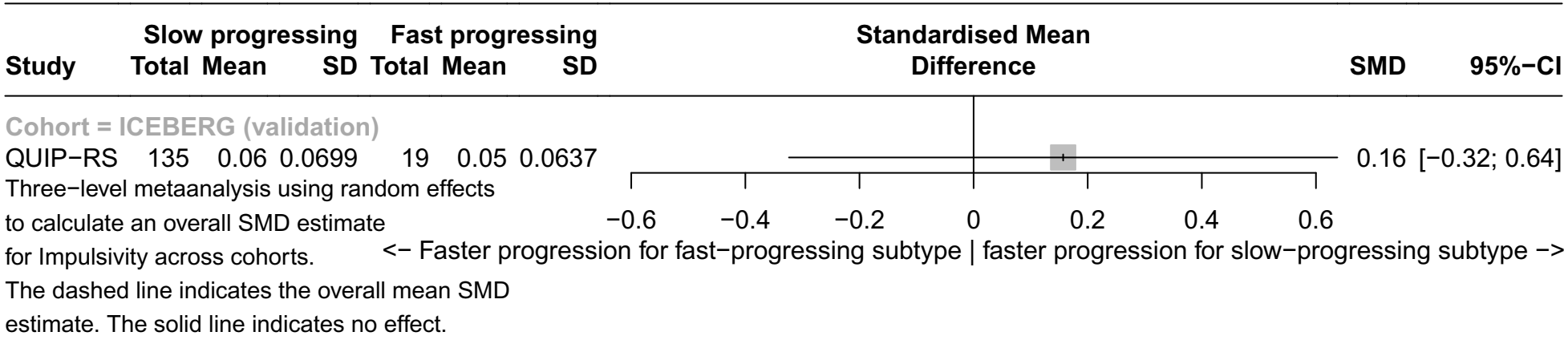
Three−level metaanalysis using random effects to calculate an overall SMD estimate for Smell across cohorts.
The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

Forest plot for progression characteristics of symptom domain Motor symptoms (validation)



Three−level metaanalysis using random effects to calculate an overall SMD estimate for Motor symptoms across cohorts. The dashed line indicates the overall mean SMD estimate. The solid line indicates no effect.

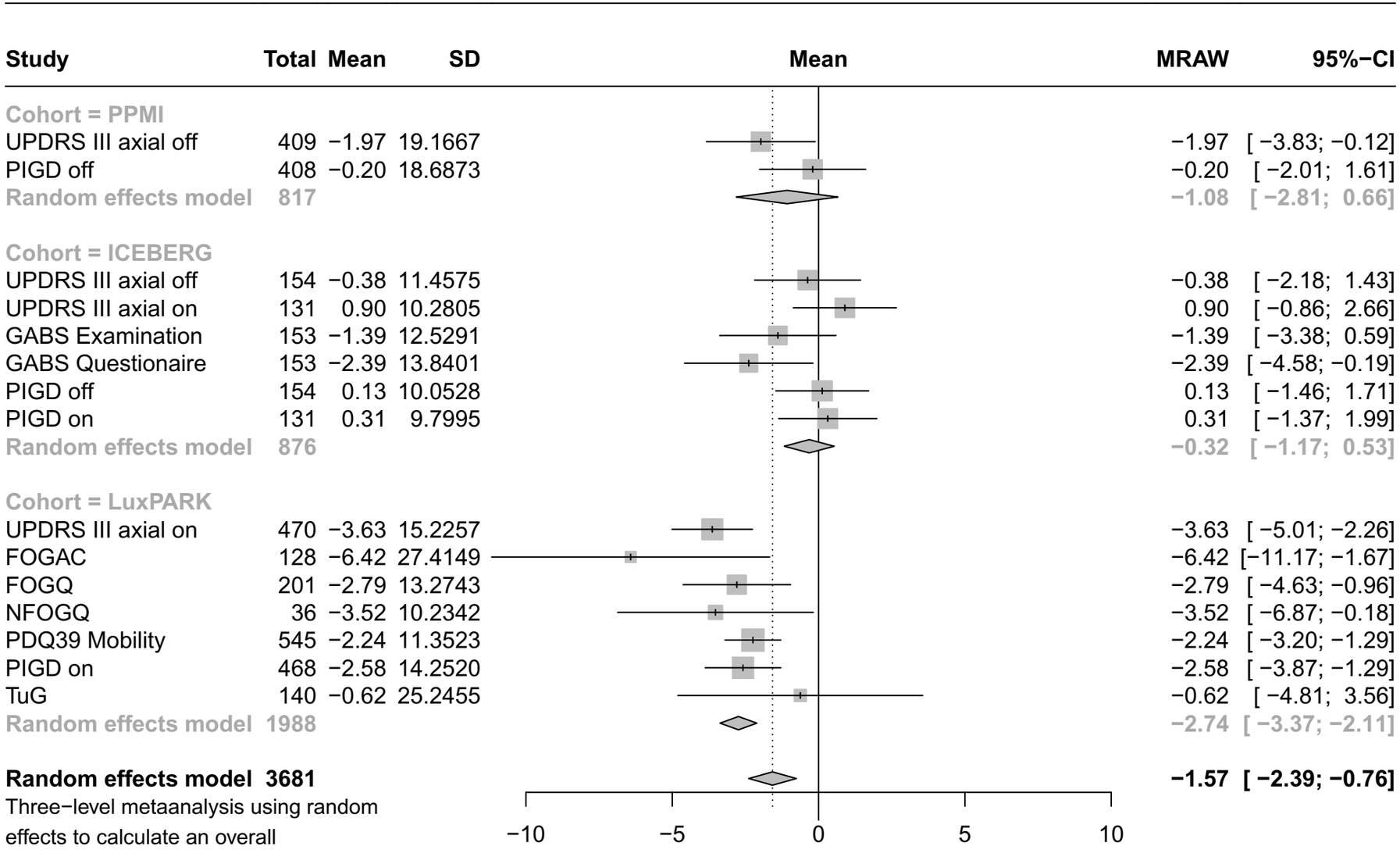
Forest plot for progression characteristics of symptom domain Impulsivity (validation)



Progression subtypes in Parkinson’s disease identified by a data driven multi cohort analysis

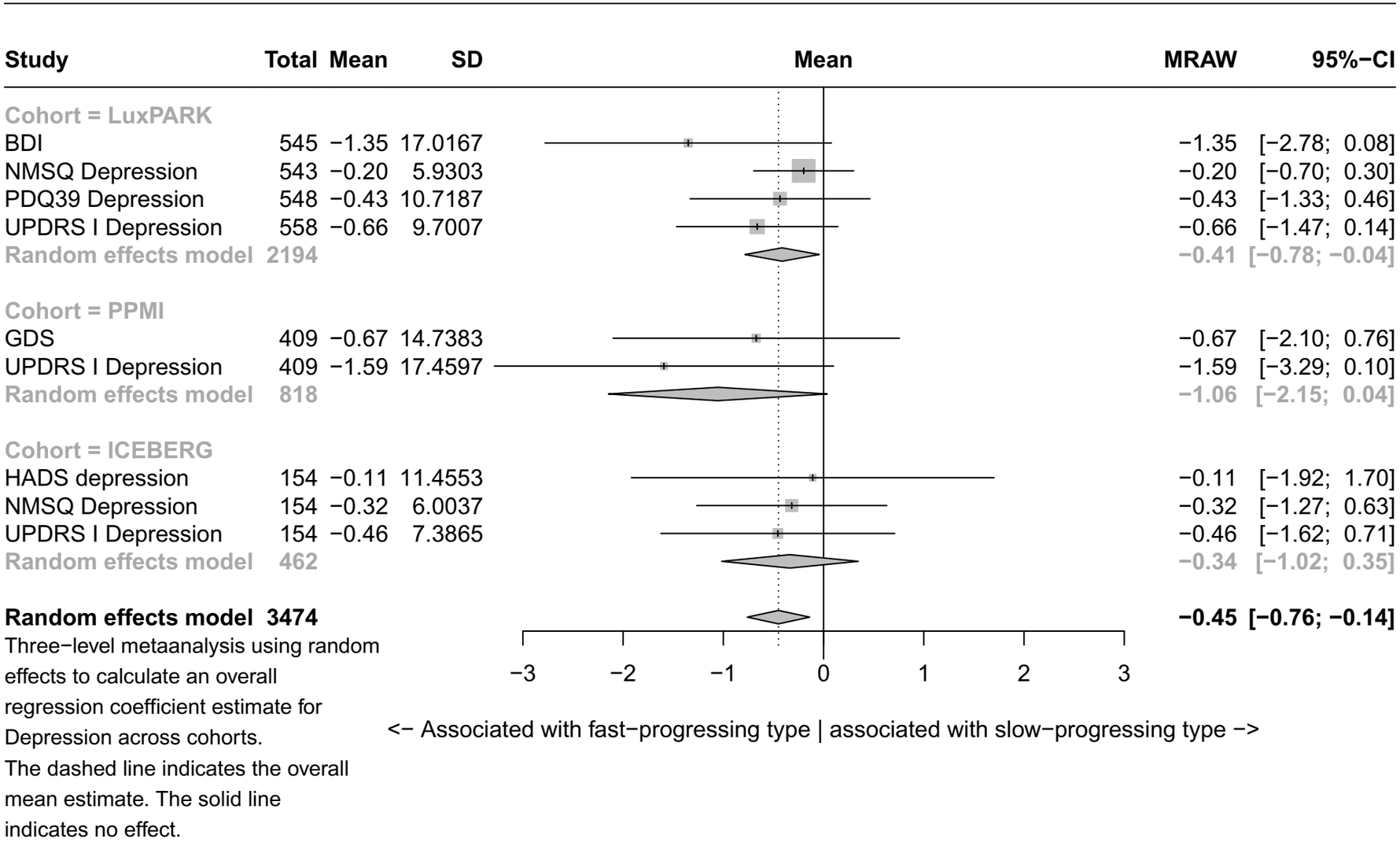
Forest plots for symptom domain baseline associations (in cohort)

Forest plot for baseline characteristics of symptom domain Axial & PIGD



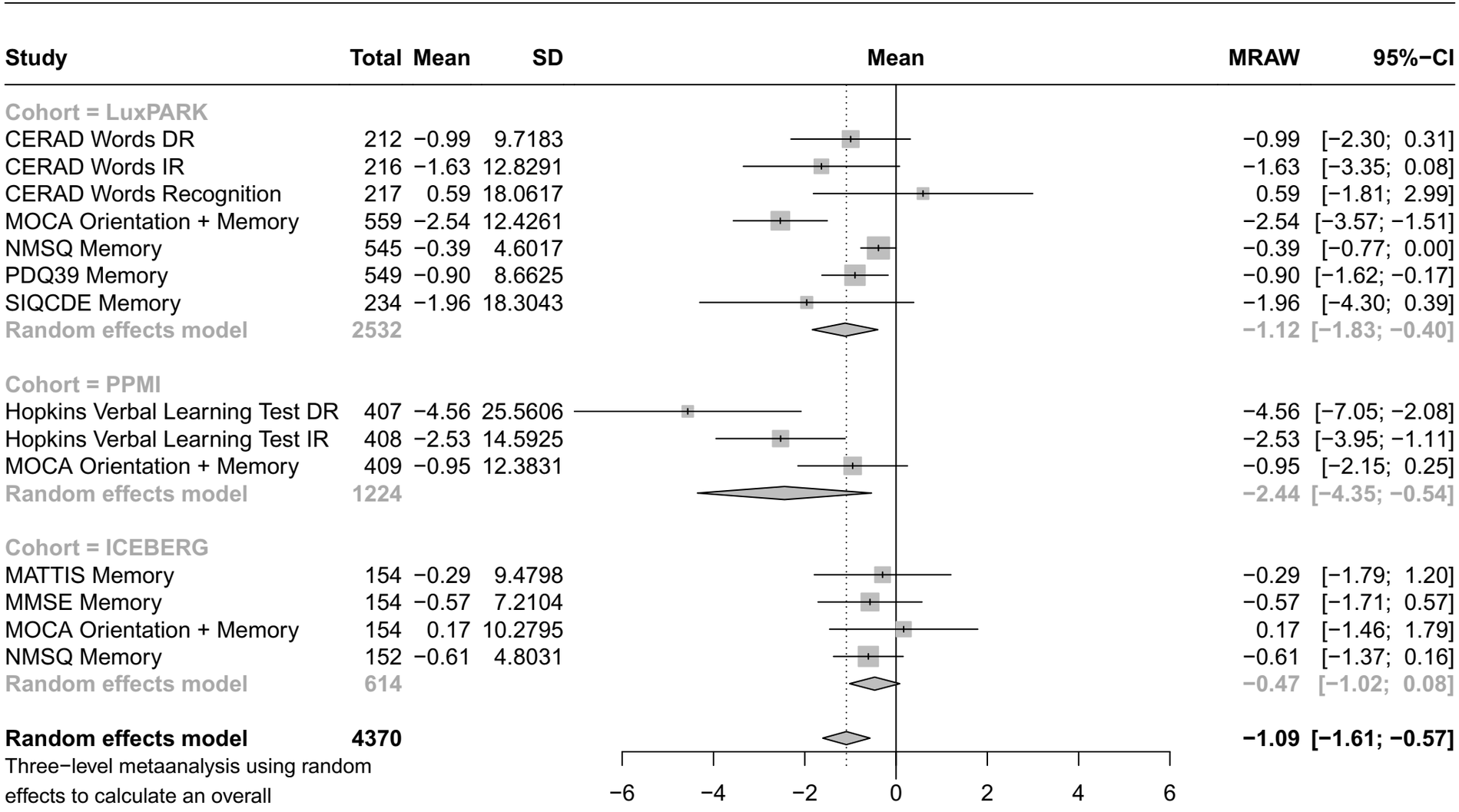
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Axial & PIGD across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Depression



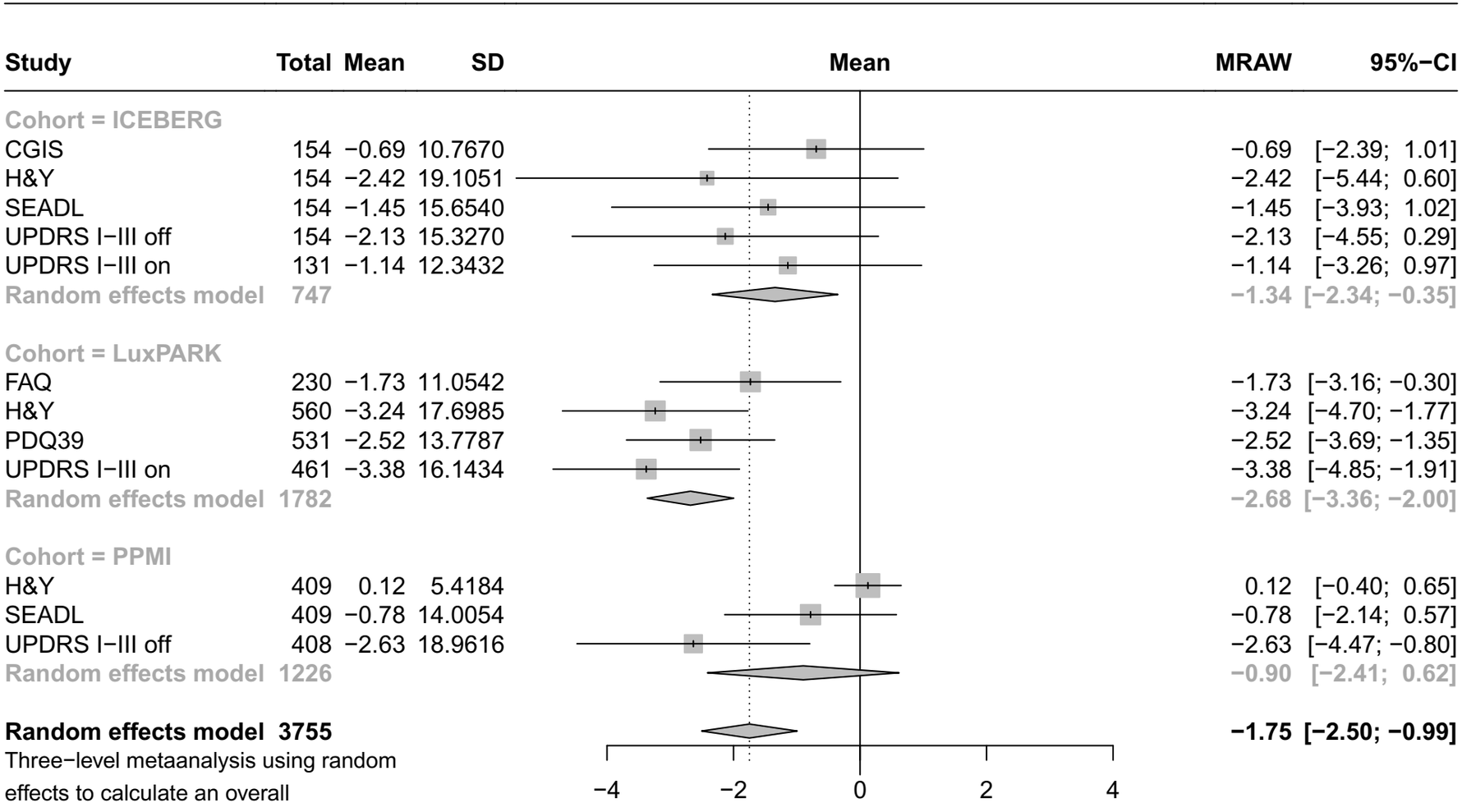
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Depression across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Memory



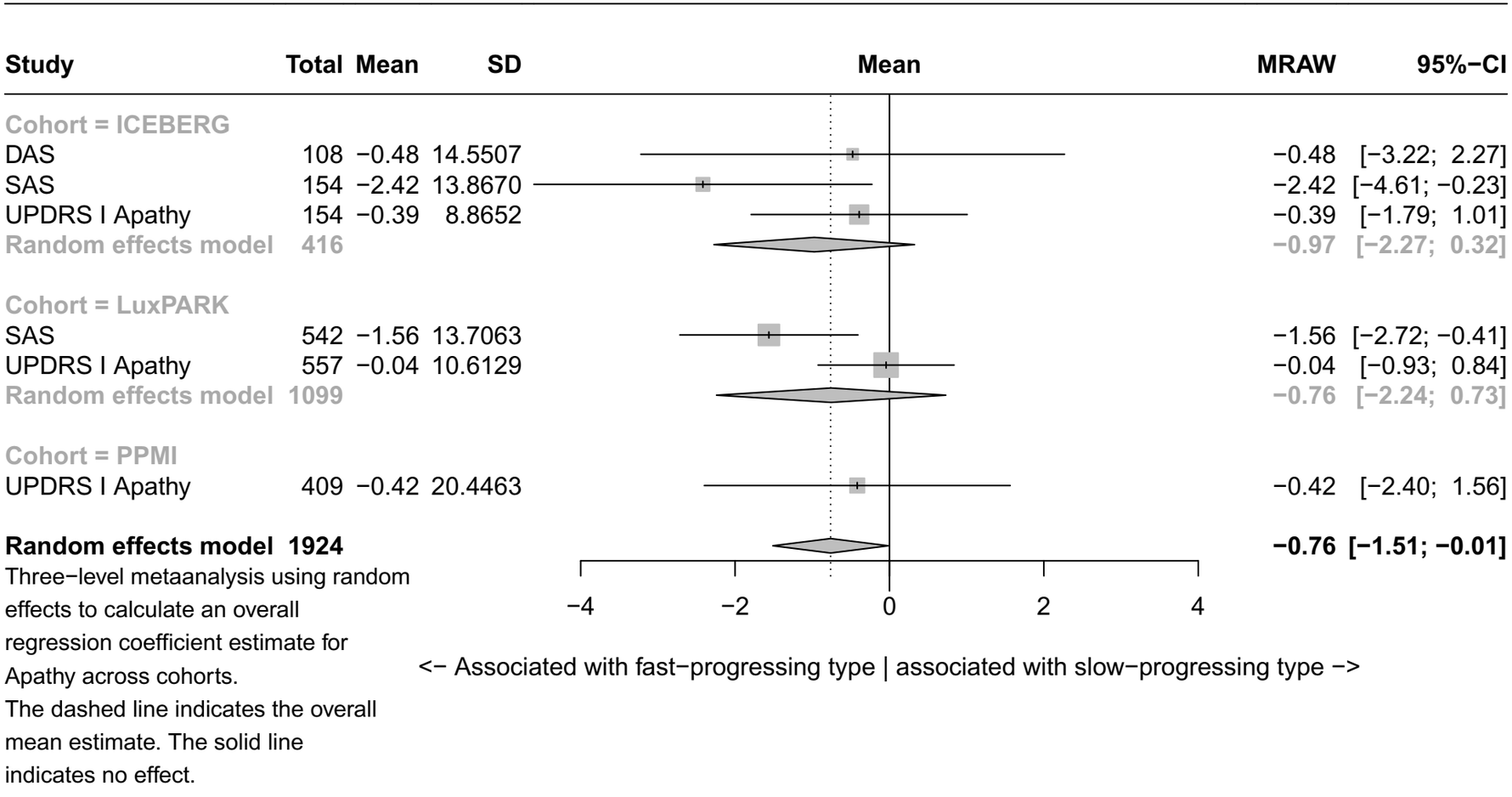
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Memory across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Overall severity



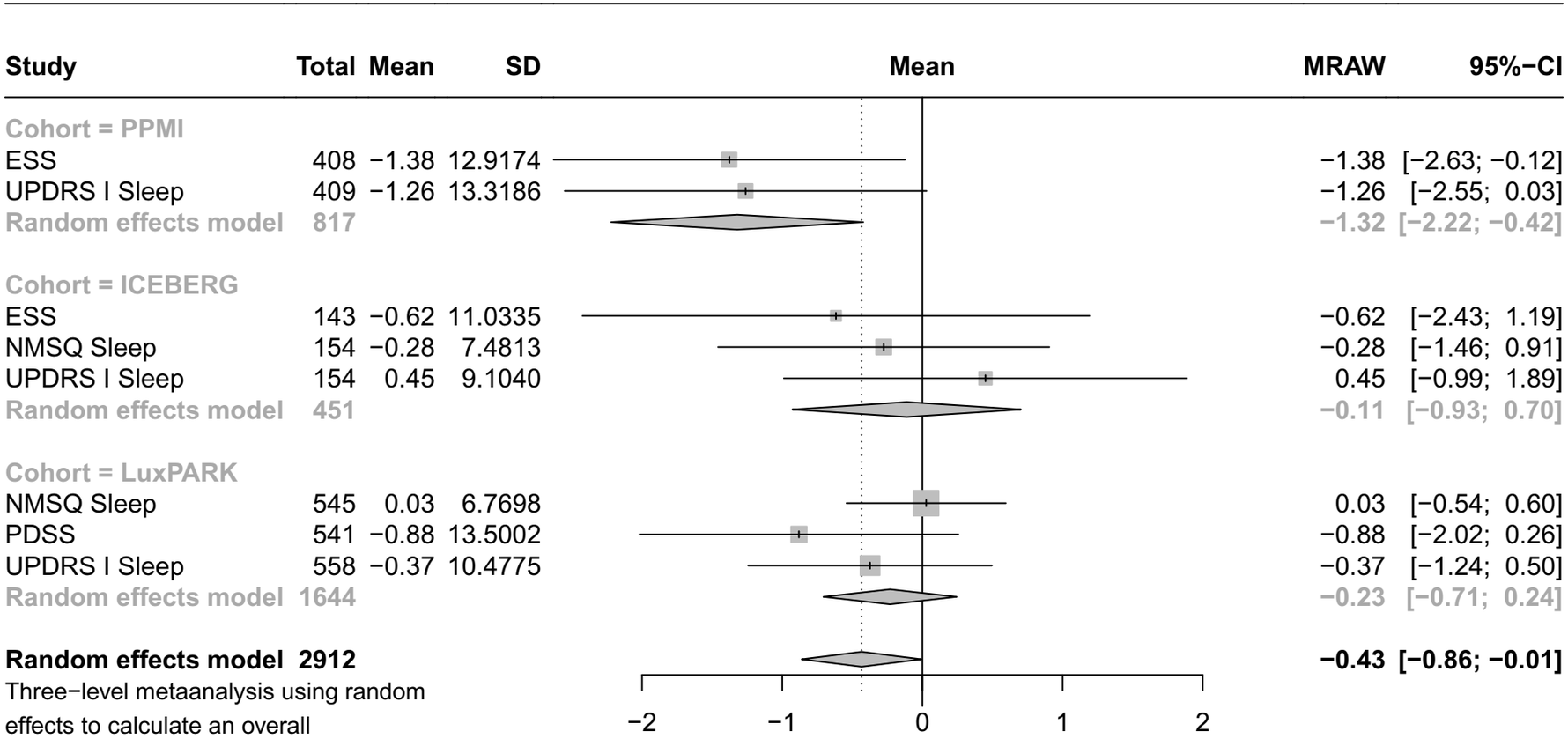
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Overall severity across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Apathy



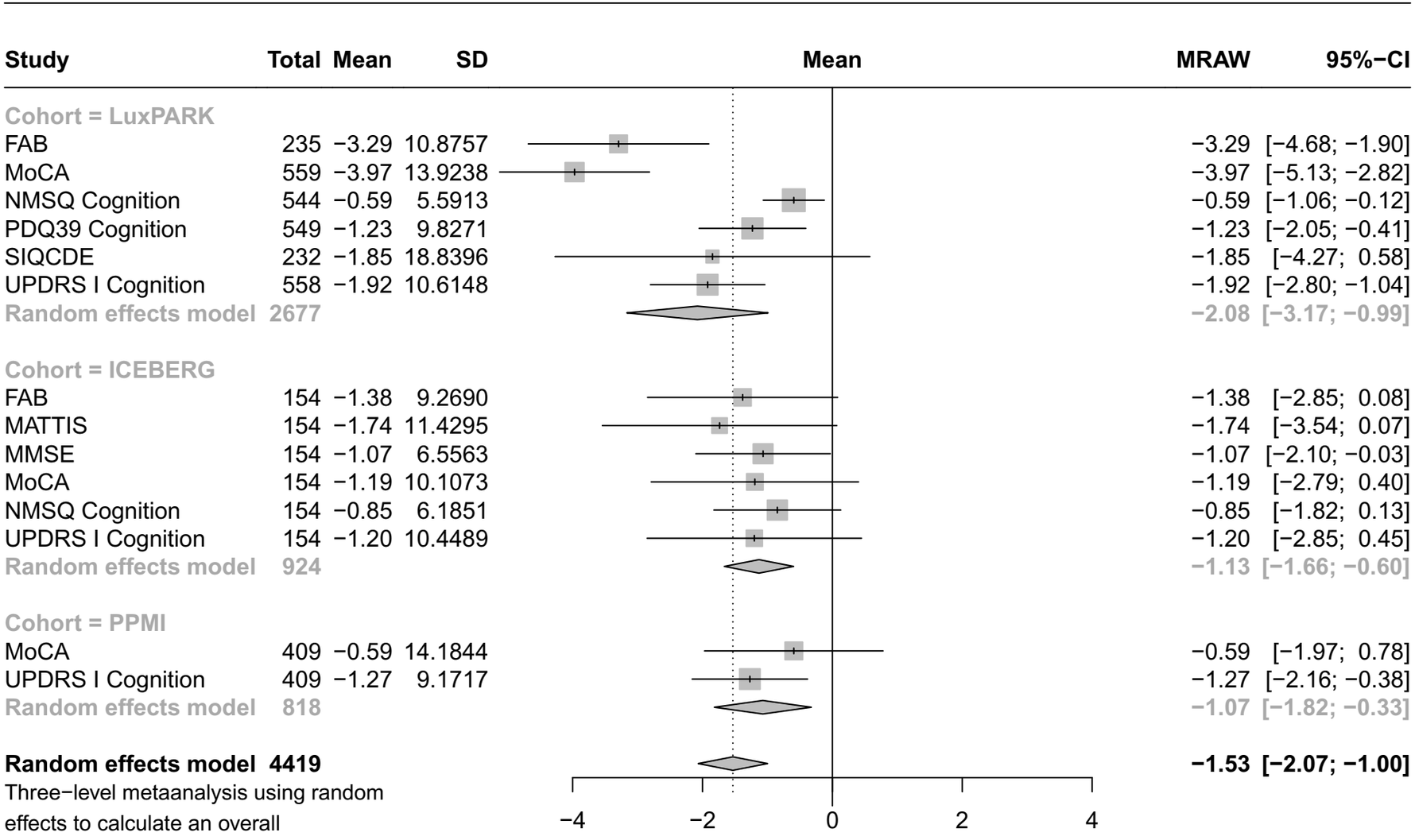
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Apathy across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Sleep



Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Sleep across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

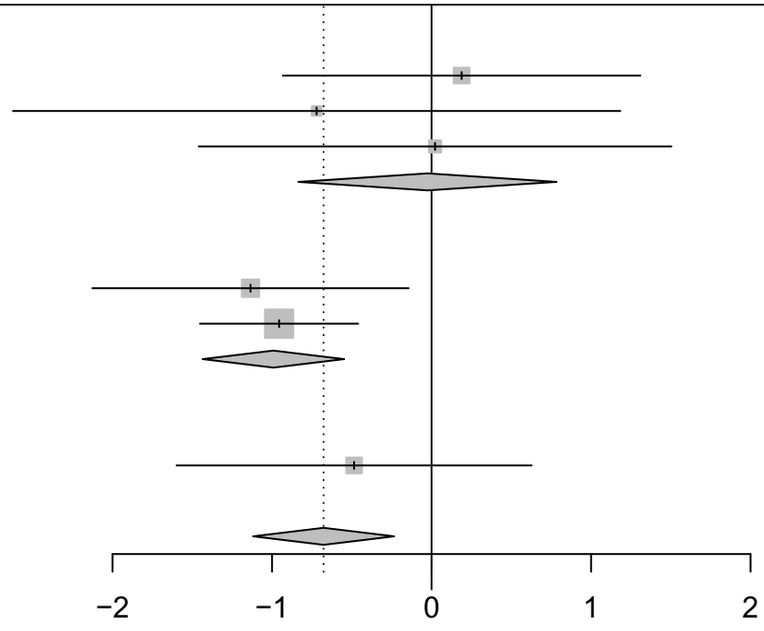
Forest plot for baseline characteristics of symptom domain Overall cognition



Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Overall cognition across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Conceptualization

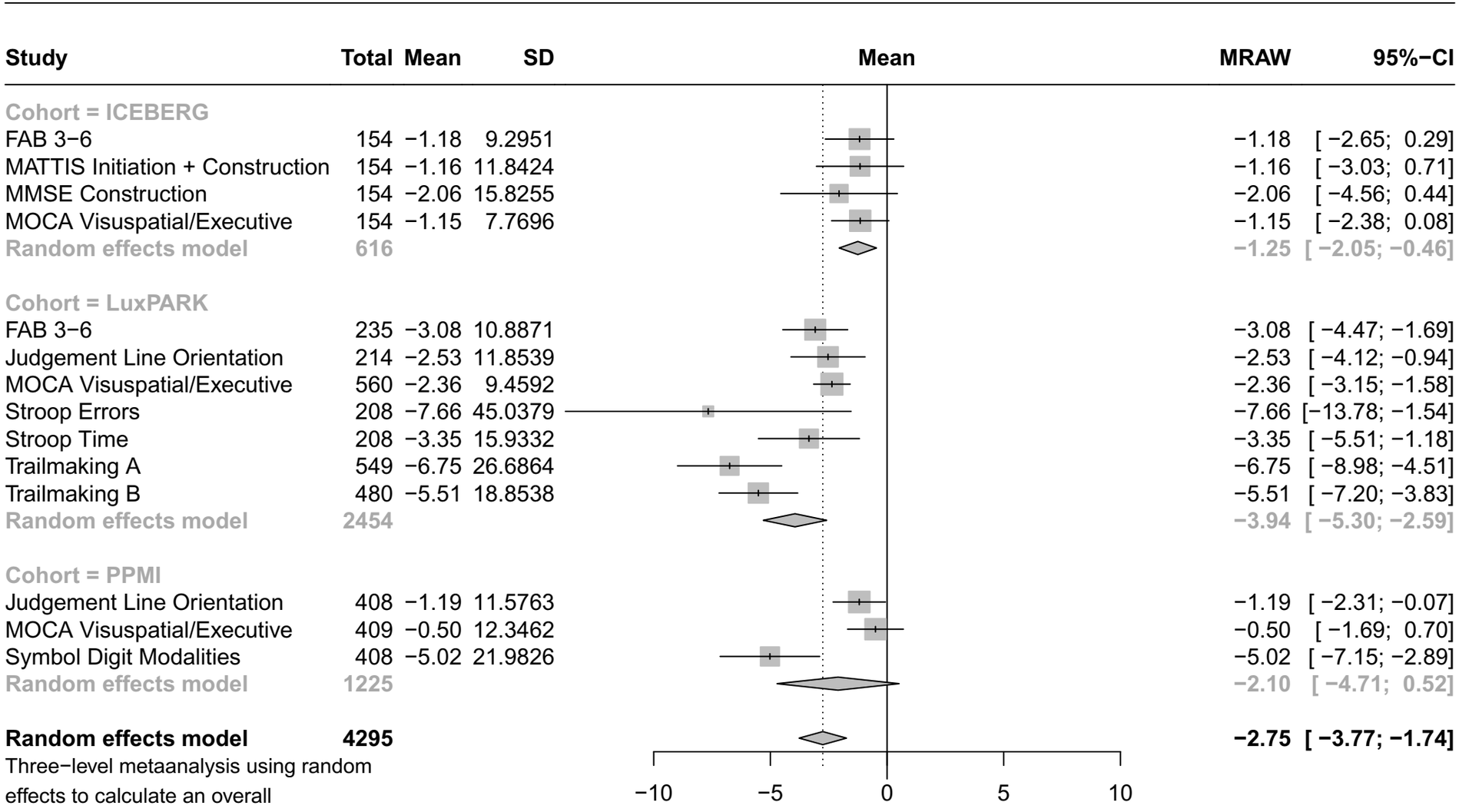
Study	Total	Mean	SD	MRAW	95%−CI
Cohort = ICEBERG					
FAB Conceptualization	154	0.19	7.0980	0.19	[−0.93; 1.31]
MATTIS Conzeptualization	154	−0.72	12.0570	−0.72	[−2.62; 1.18]
MOCA Abstraction	154	0.02	9.3849	0.02	[−1.46; 1.50]
Random effects model	**462**			**−0.03**	**[−0.83; 0.78]**
Cohort = LuxPARK					
FAB Conceptualization	236	−1.14	7.7740	−1.14	[−2.13; −0.14]
MOCA Abstraction	559	−0.96	5.9852	−0.96	[−1.45; −0.46]
Random effects model	**795**			**−0.99**	**[−1.44; −0.55]**
Cohort = PPMI					
MOCA Abstraction	409	−0.49	11.4891	−0.49	[−1.60; 0.63]
Random effects model	**1666**			**−0.68**	**[−1.12; −0.23]**



<− Associated with fast−progressing type | associated with slow−progressing type −>

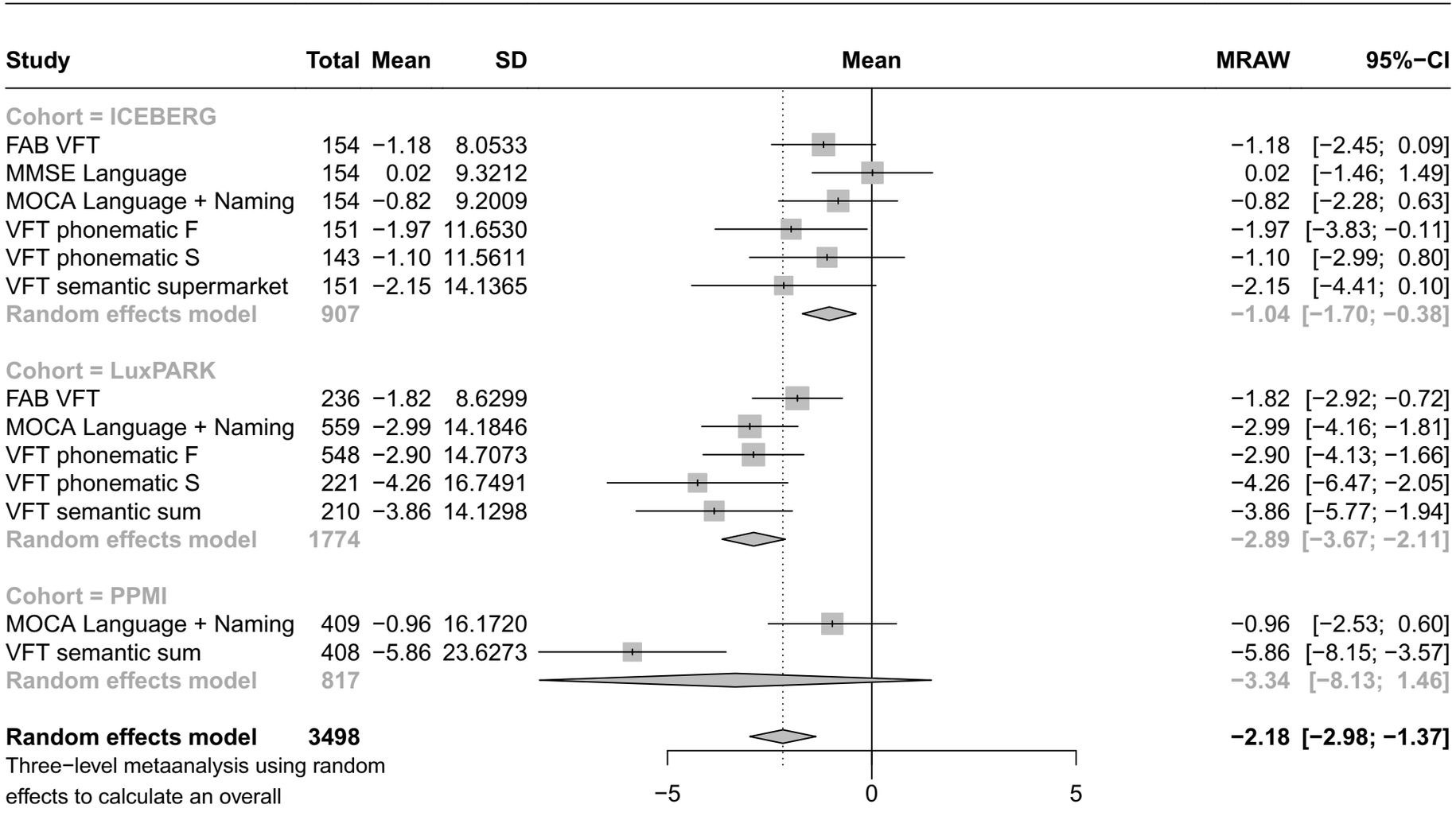
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Conceptualization across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Visuo−executive



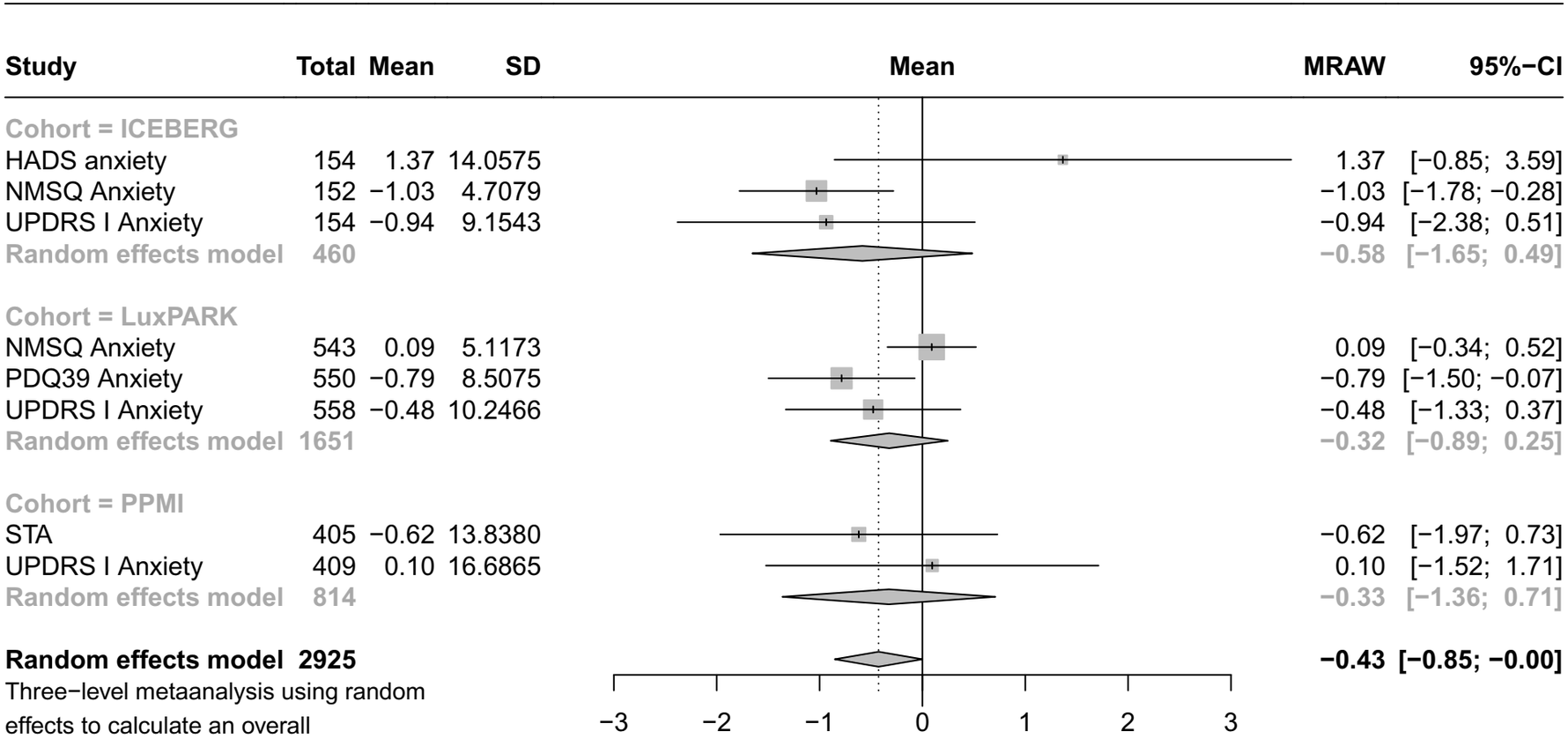
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Visuo−executive across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Language



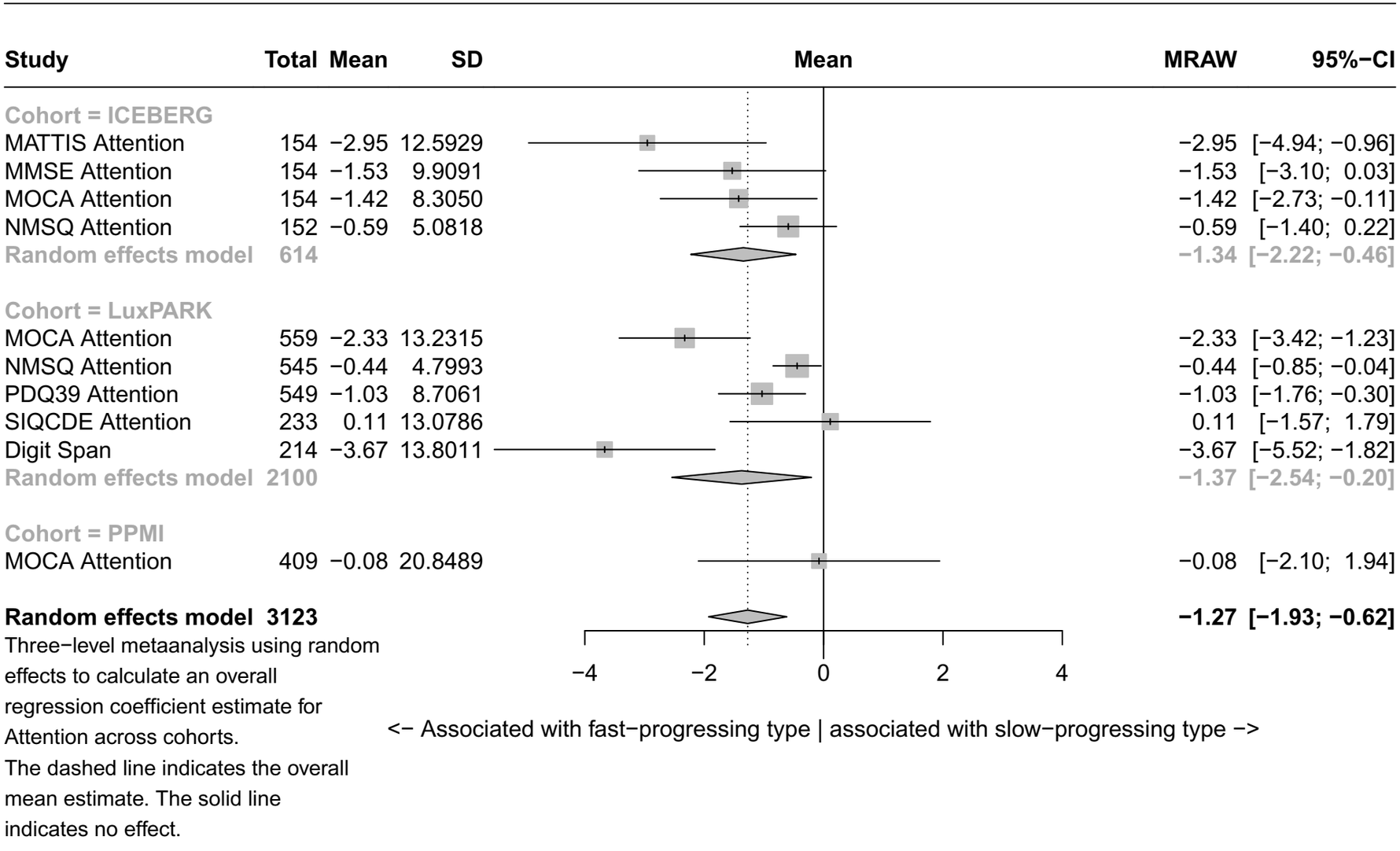
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Language across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Anxiety



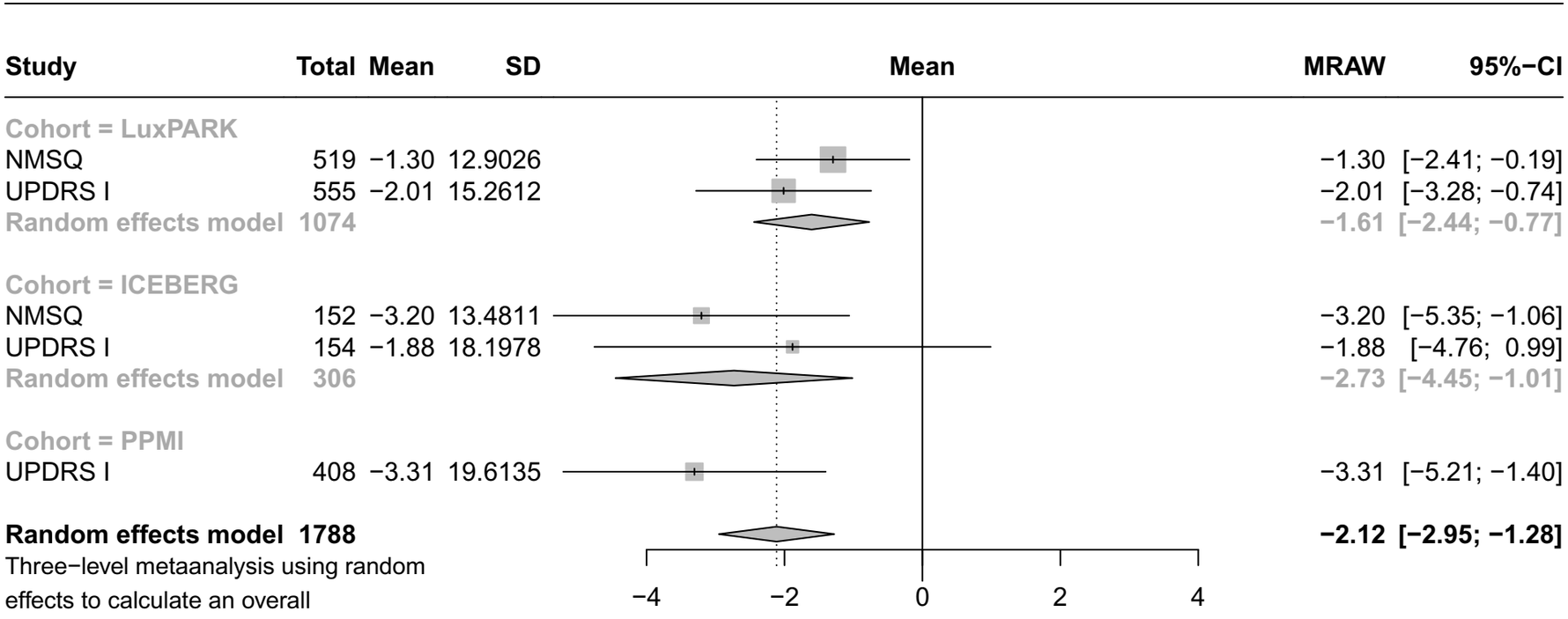
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Anxiety across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Attention



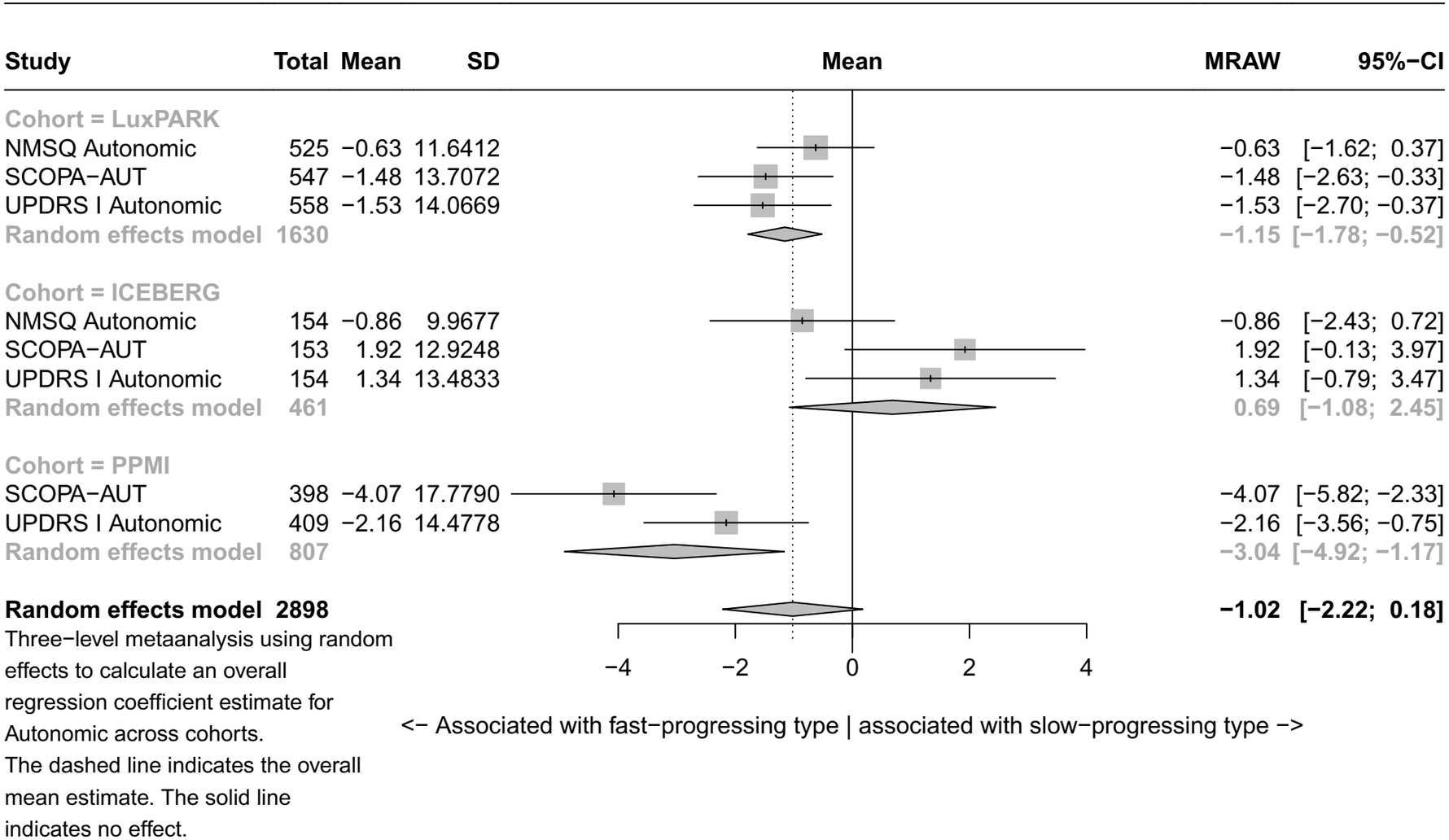
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Attention across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Non motor symptoms



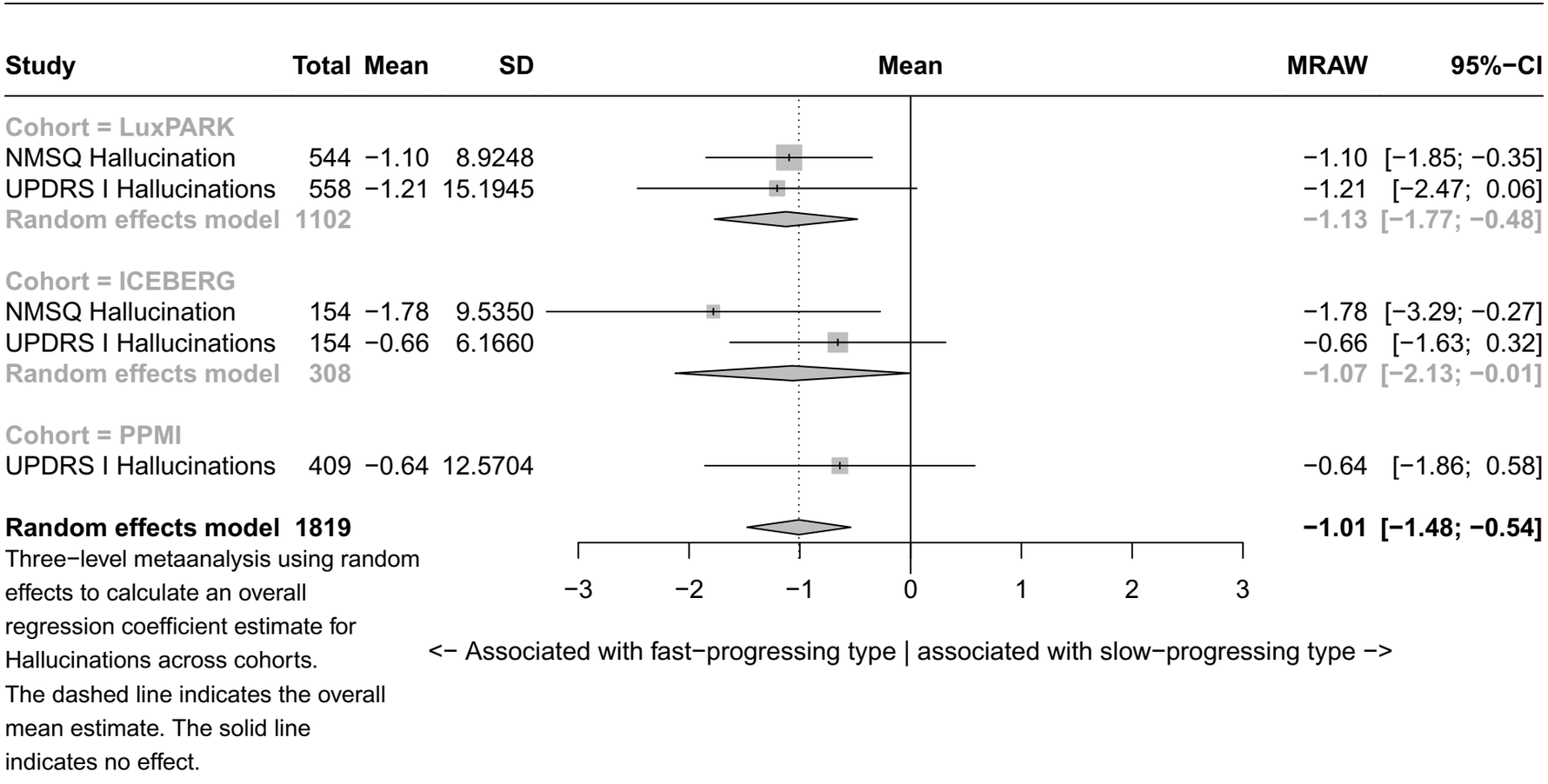
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Non motor symptoms across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Autonomic



Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Autonomic across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

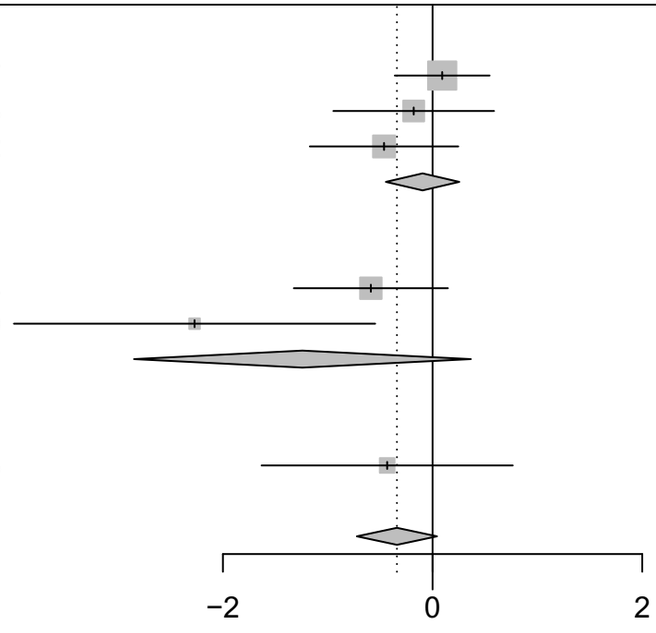
Forest plot for baseline characteristics of symptom domain Hallucinations



Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Hallucinations across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Pain

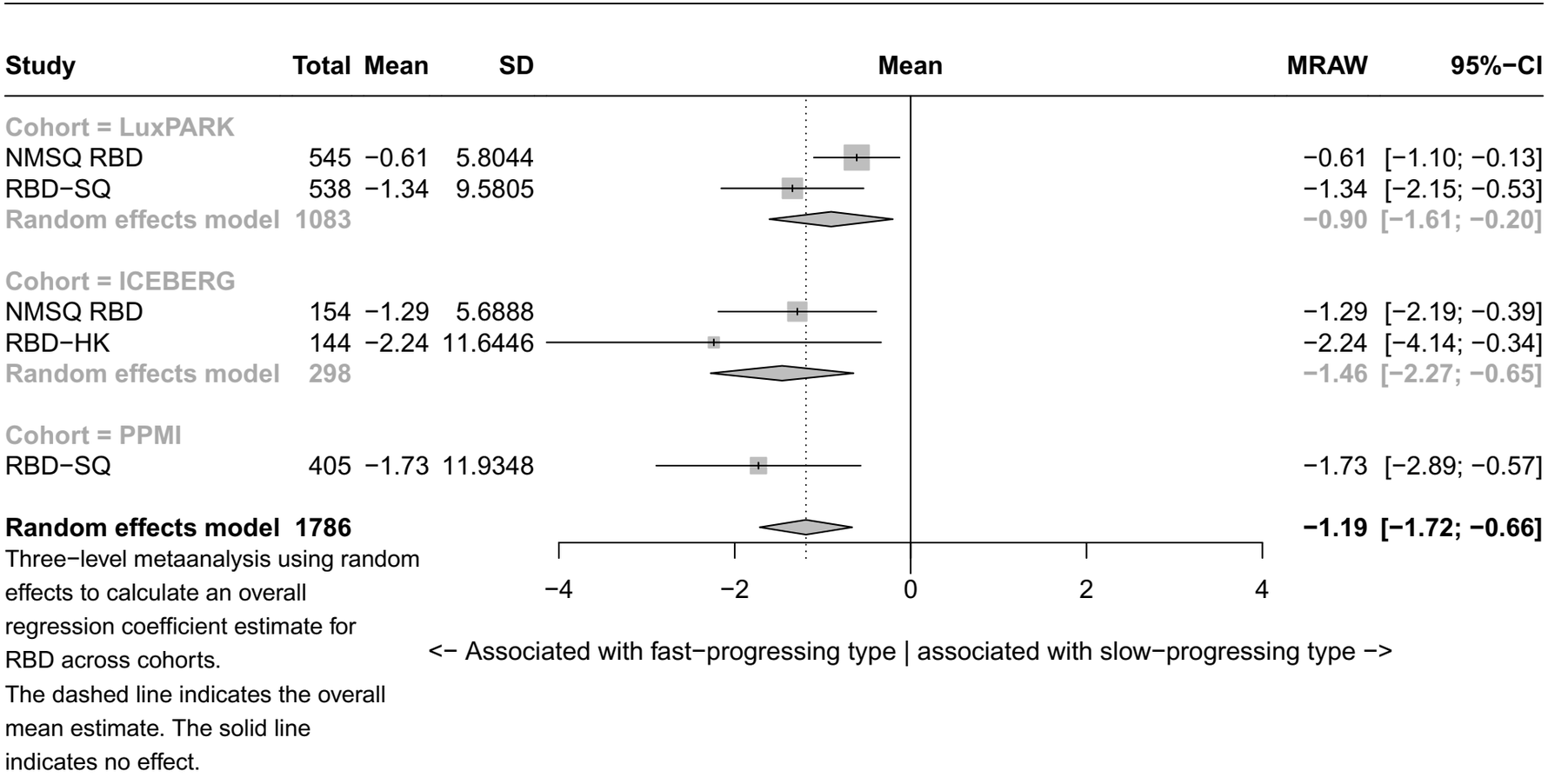
Study	Total	Mean	SD	MRAW	95%−CI
Cohort = LuxPARK					
NMSQ Pain	546	0.09	5.3917	0.09	[−0.36; 0.54]
PDQ39 Pain	547	−0.18	9.1475	−0.18	[−0.95; 0.59]
UPDRS I Pain	558	−0.46	8.5352	−0.46	[−1.17; 0.25]
Random effects model	**1651**			**−0.10**	**[−0.45; 0.26]**
Cohort = ICEBERG					
NMSQ Pain	152	−0.59	4.6244	−0.59	[−1.32; 0.15]
UPDRS I Pain	154	−2.27	10.9099	−2.27	[−3.99; −0.55]
Random effects model	**306**			**−1.24**	**[−2.85; 0.36]**
Cohort = PPMI					
UPDRS I Pain	408	−0.43	12.3475	−0.43	[−1.63; 0.77]
Random effects model	**2365**			**−0.34**	**[−0.72; 0.04]**



<− Associated with fast−progressing type | associated with slow−progressing type −>

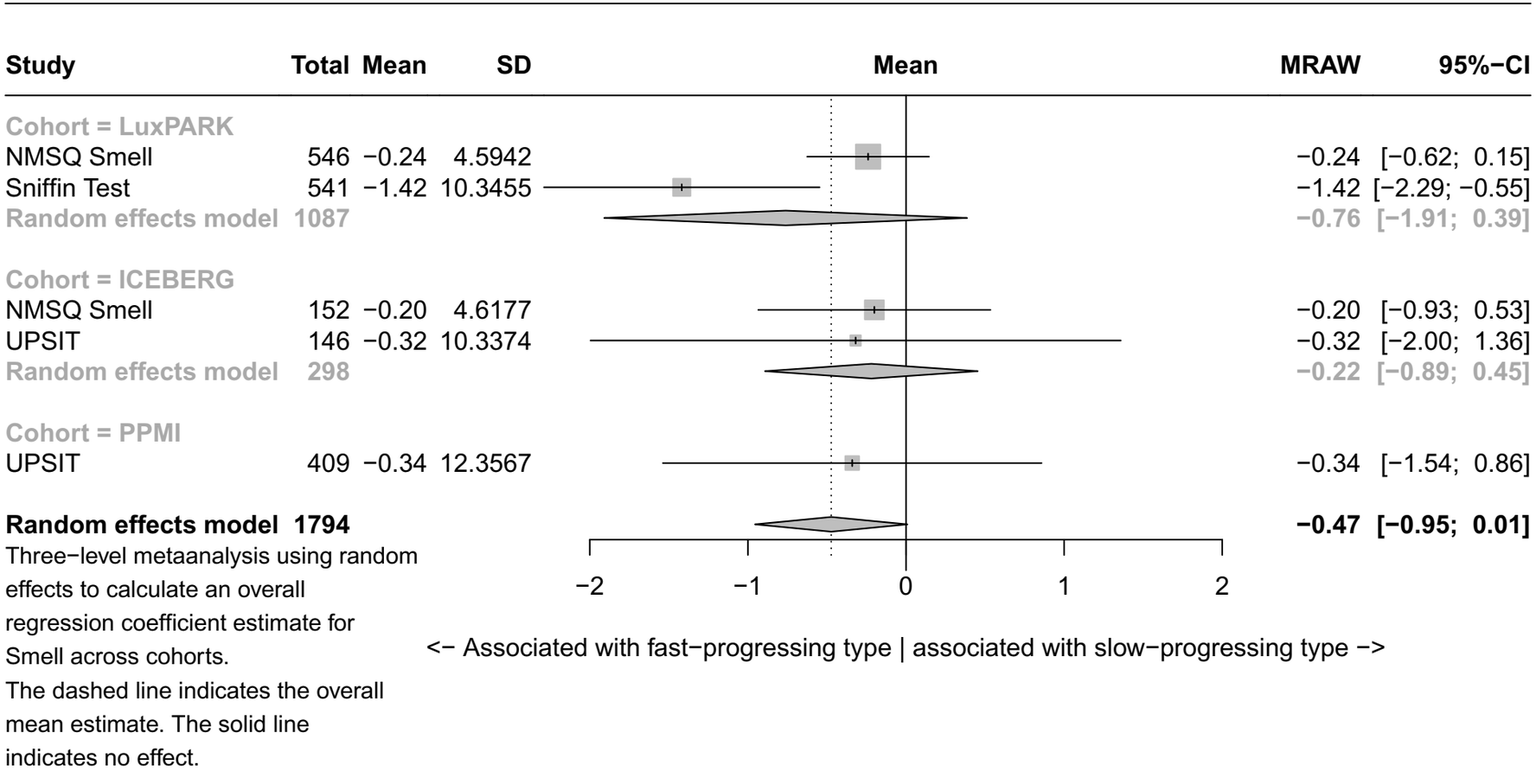
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Pain across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain RBD



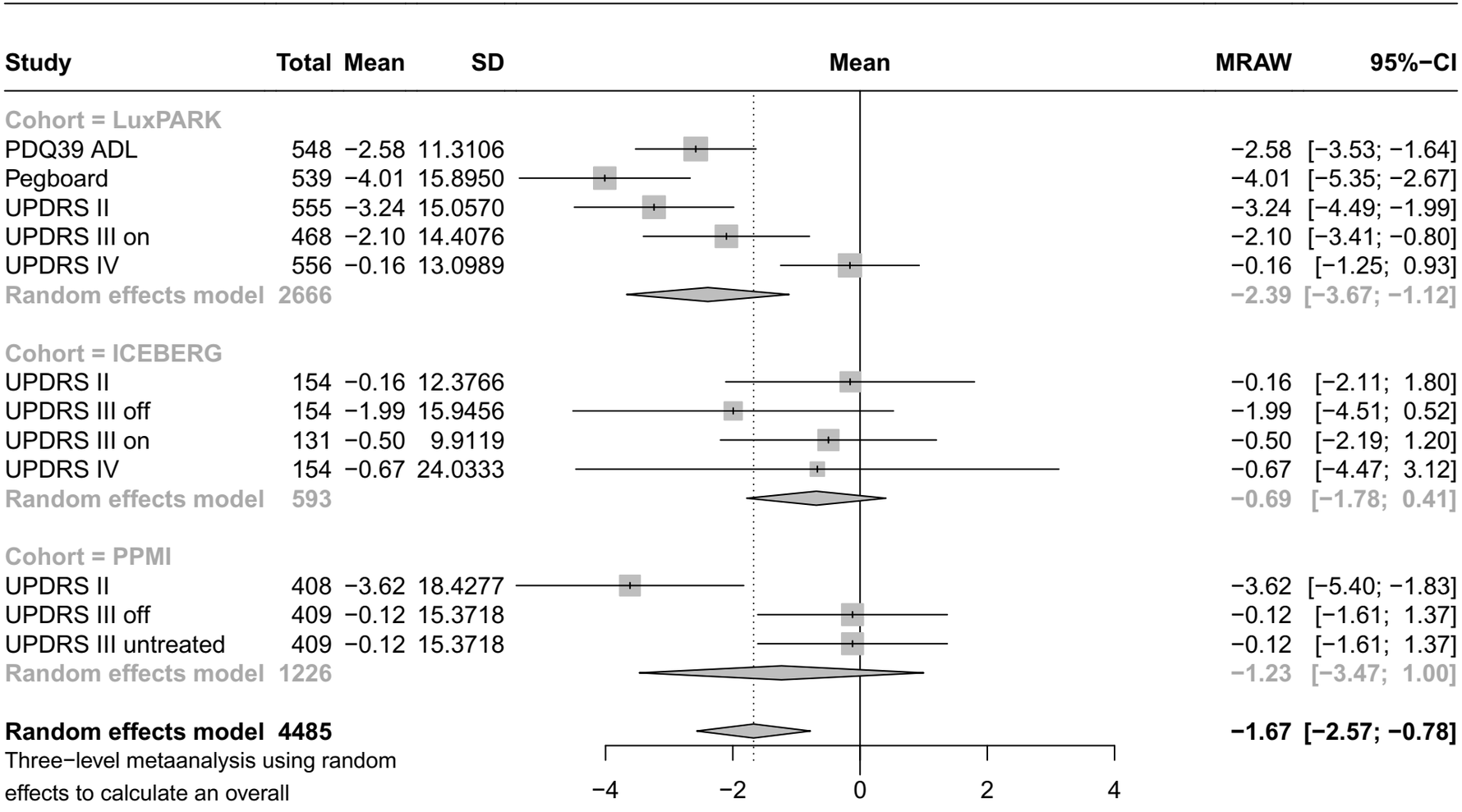
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for RBD across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Smell



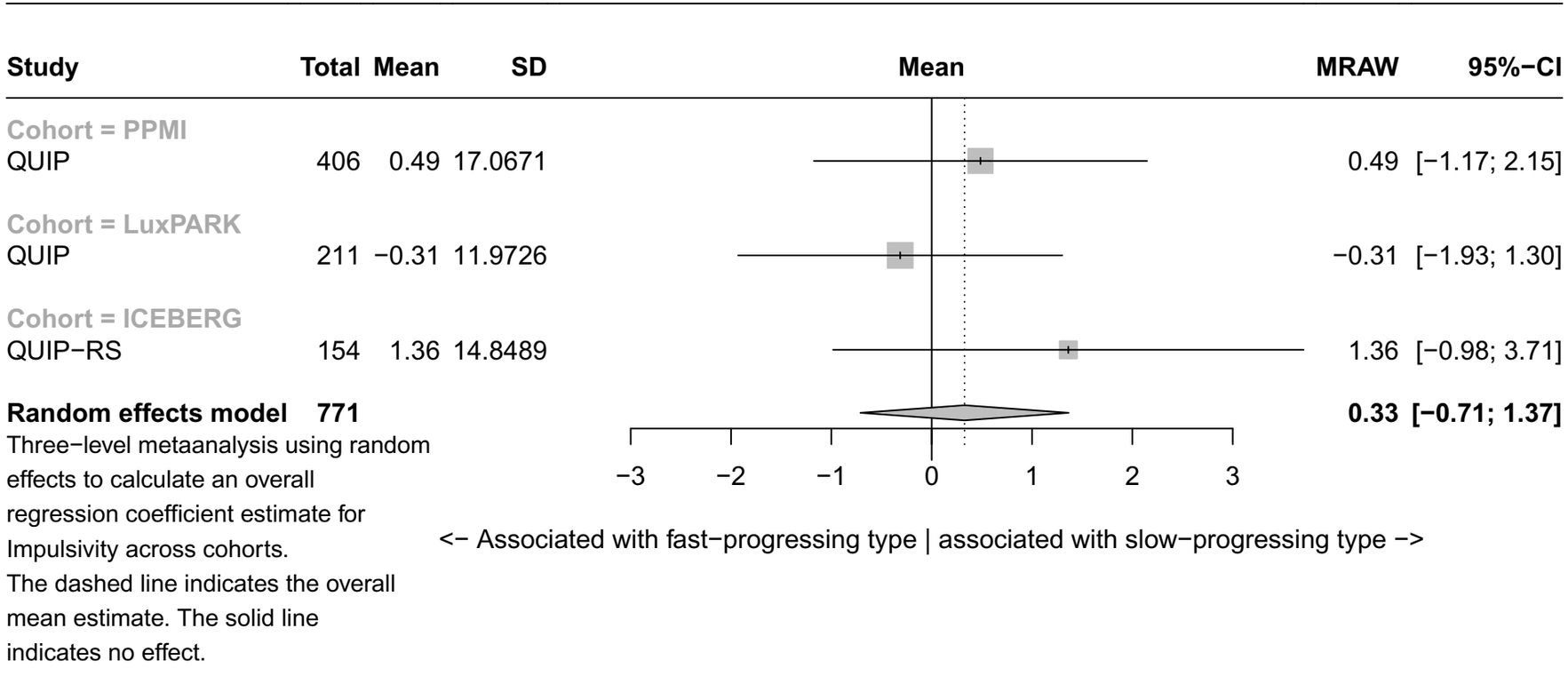
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Smell across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Motor symptoms



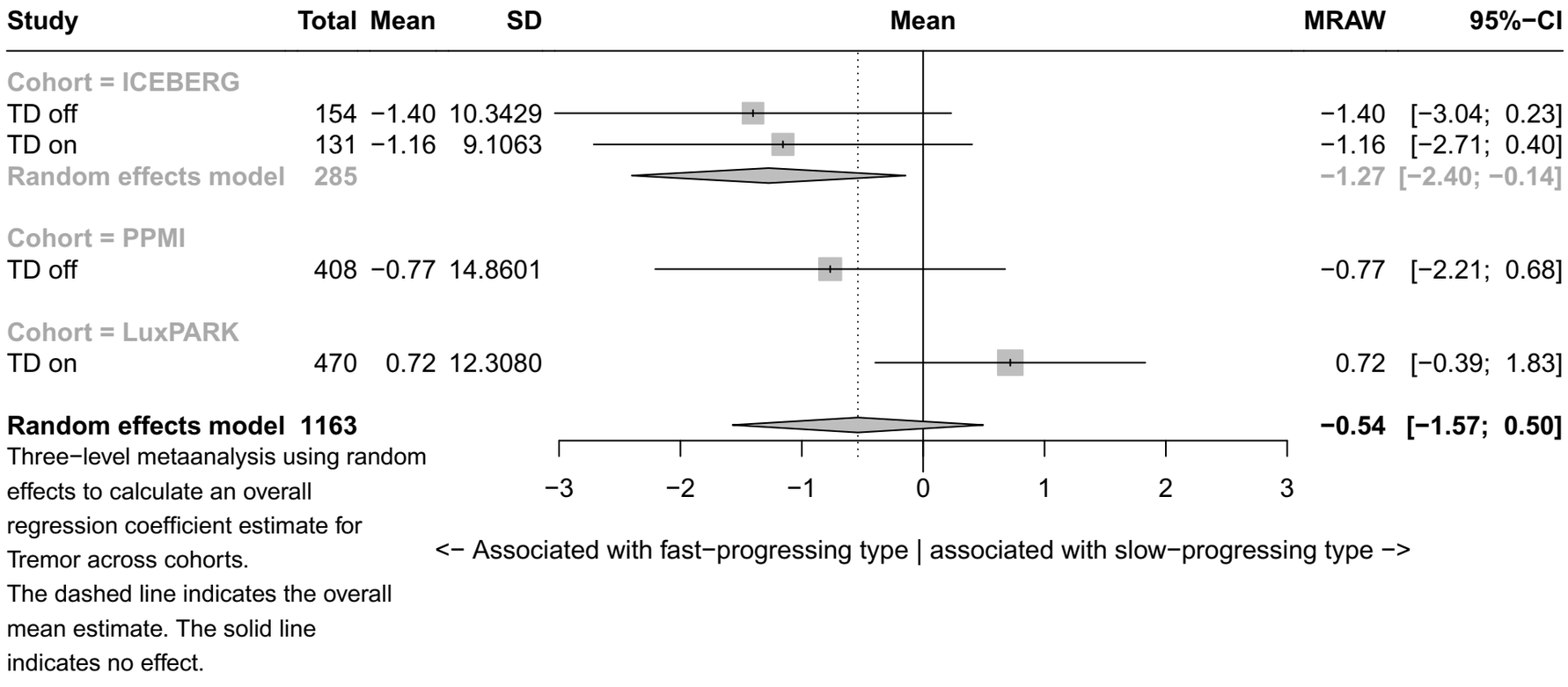
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Motor symptoms across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Impulsivity



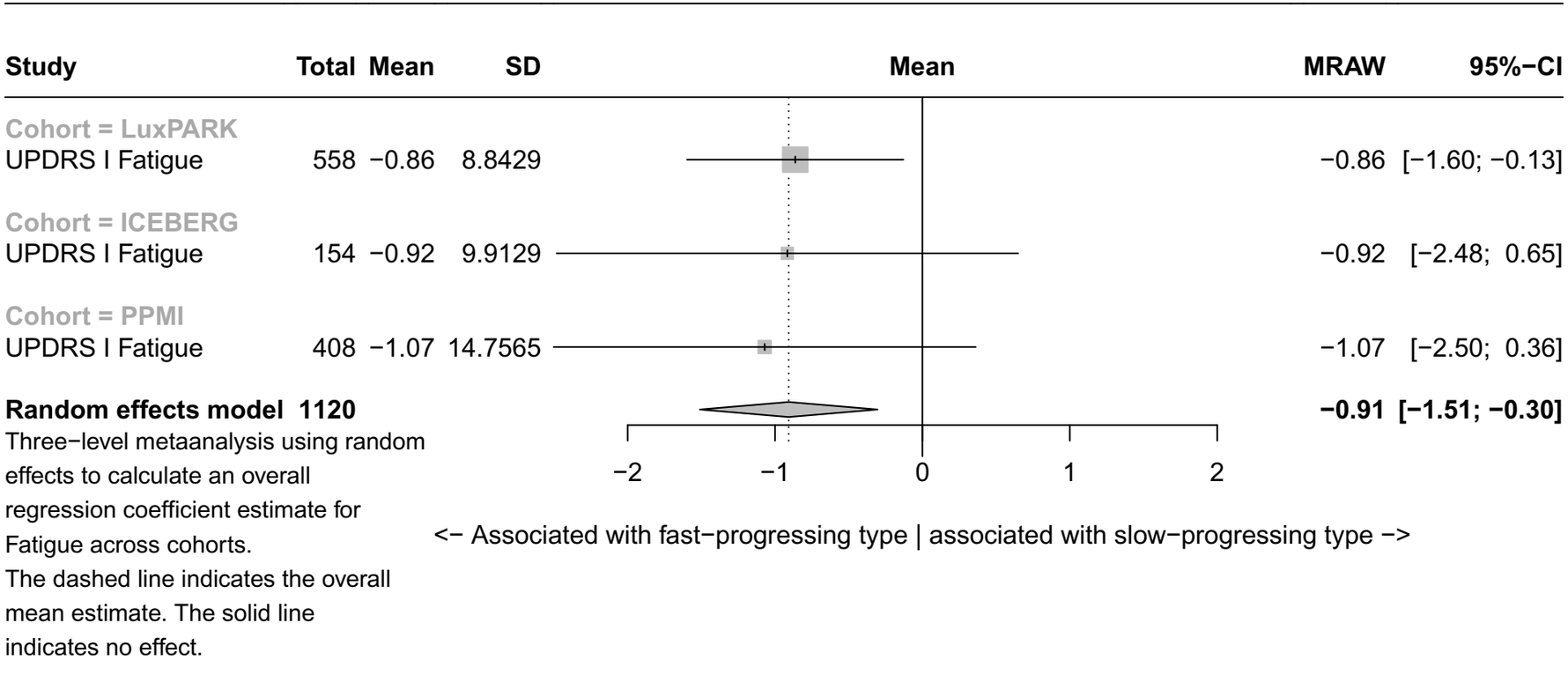
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Impulsivity across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Tremor



Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Tremor across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Fatigue

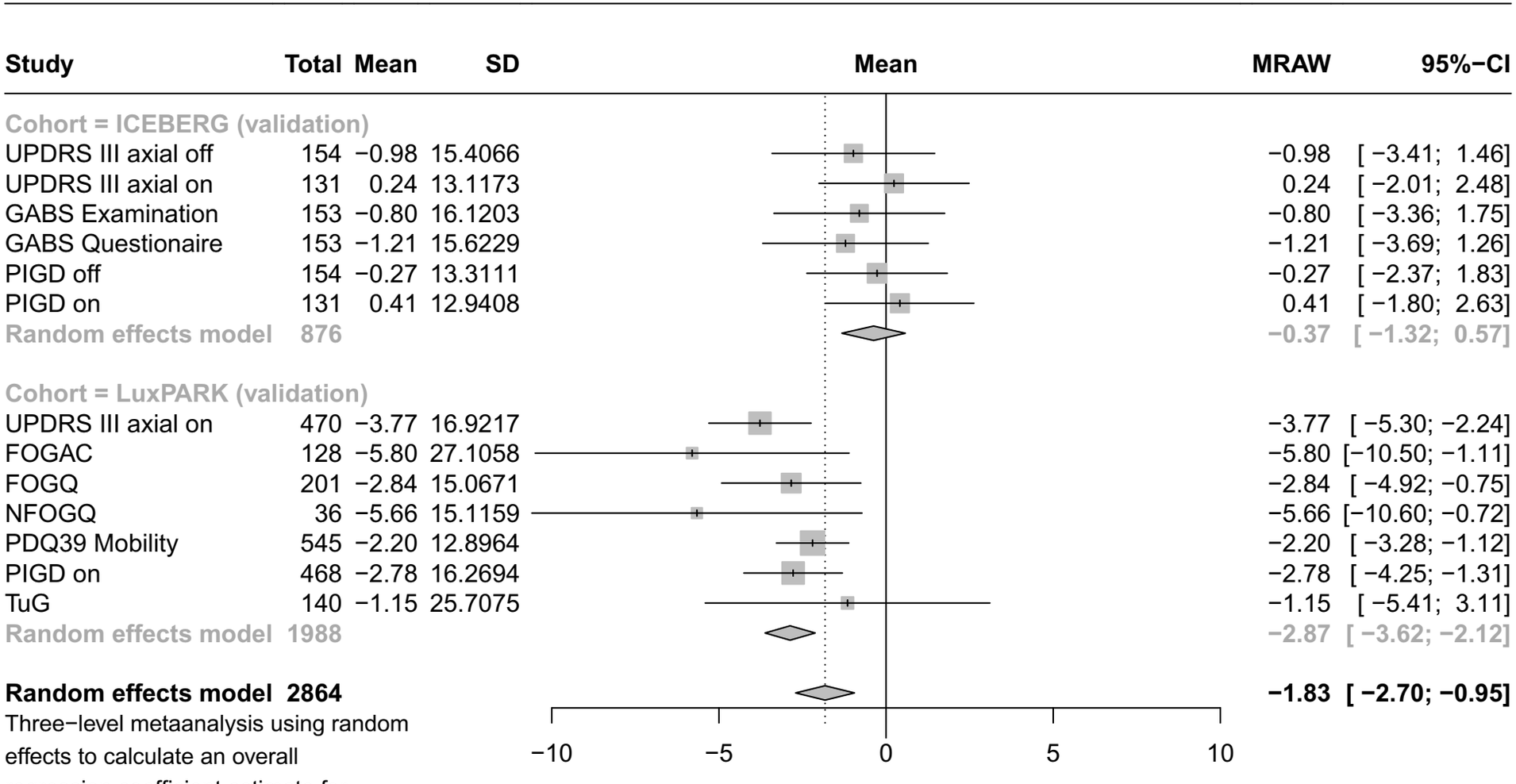


Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Fatigue across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Progression subtypes in Parkinson’s disease identified by a data driven multi cohort analysis

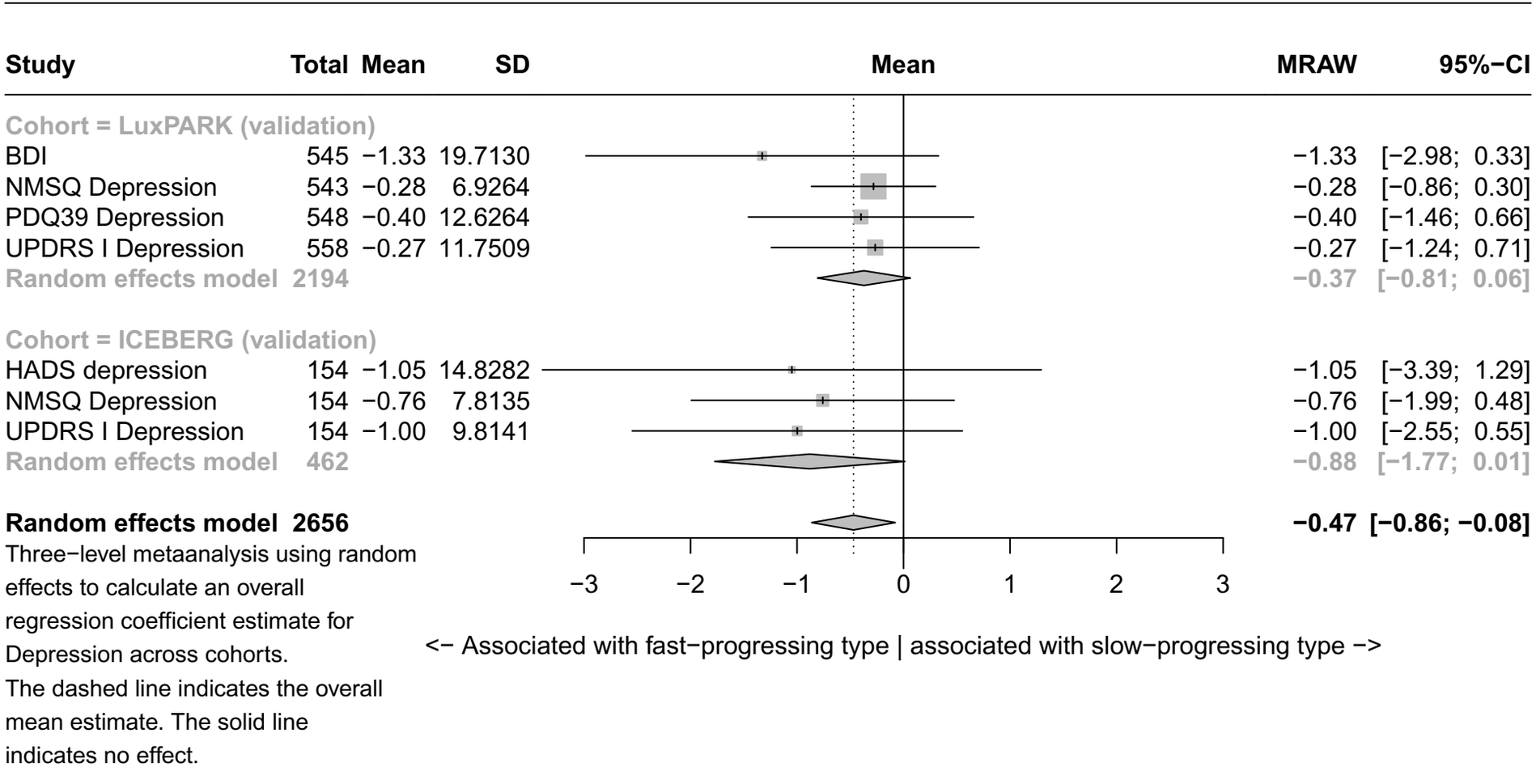
Forest plots for symptom domain baseline associations (cross-cohort validation)

Forest plot for baseline characteristics of symptom domain Axial & PIGD (validation)



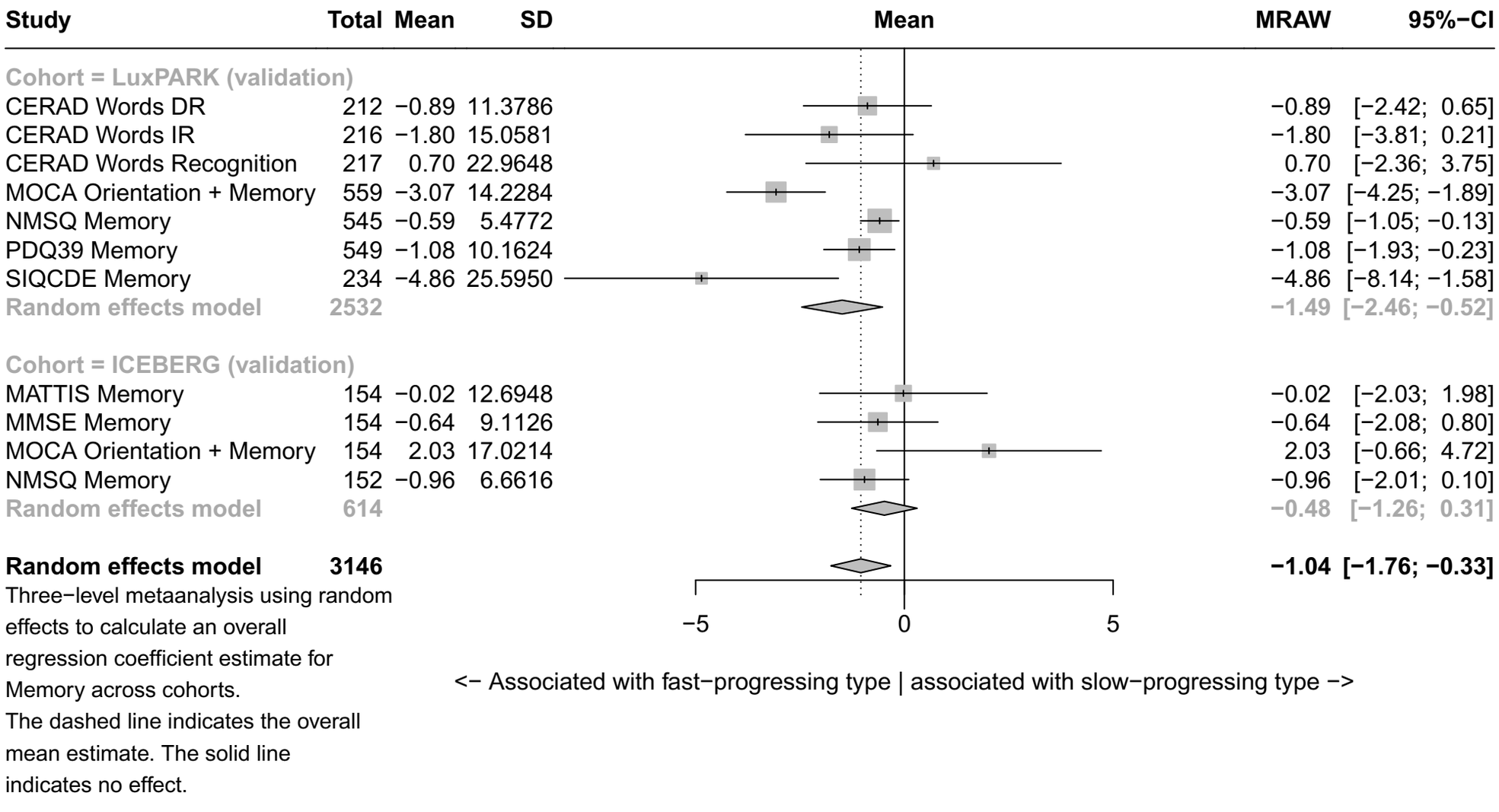
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Axial & PIGD across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Depression (validation)



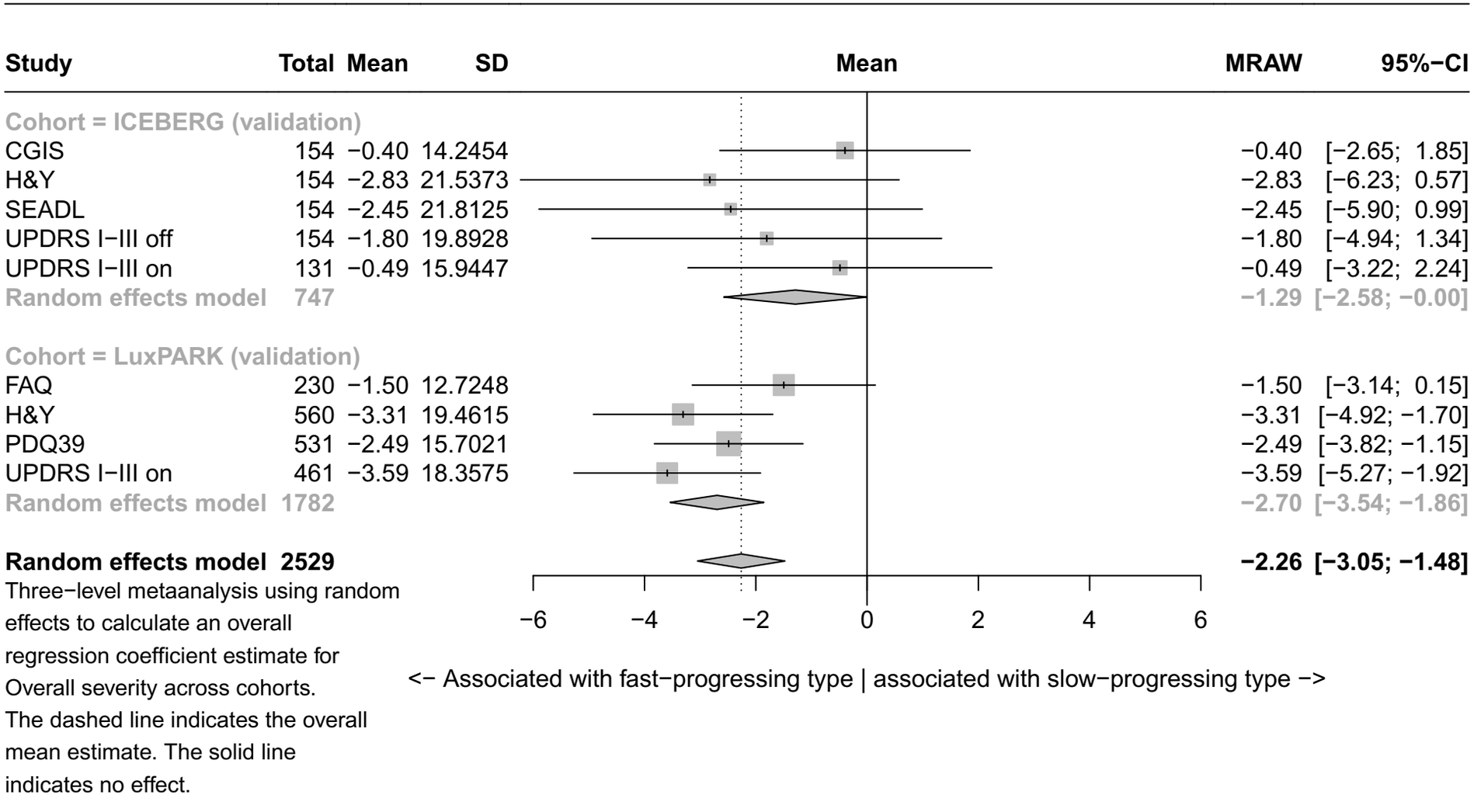
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Depression across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Memory (validation)



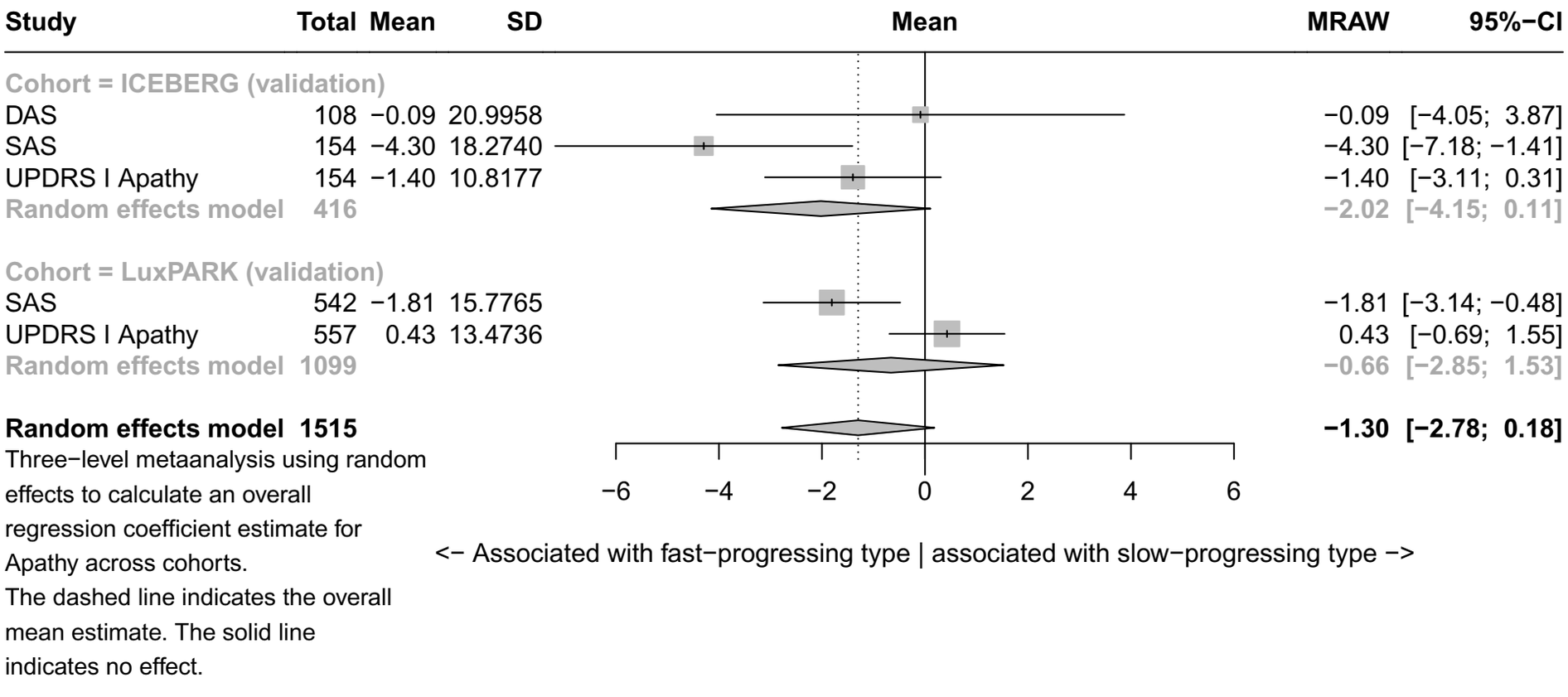
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Memory across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Overall severity (validation)



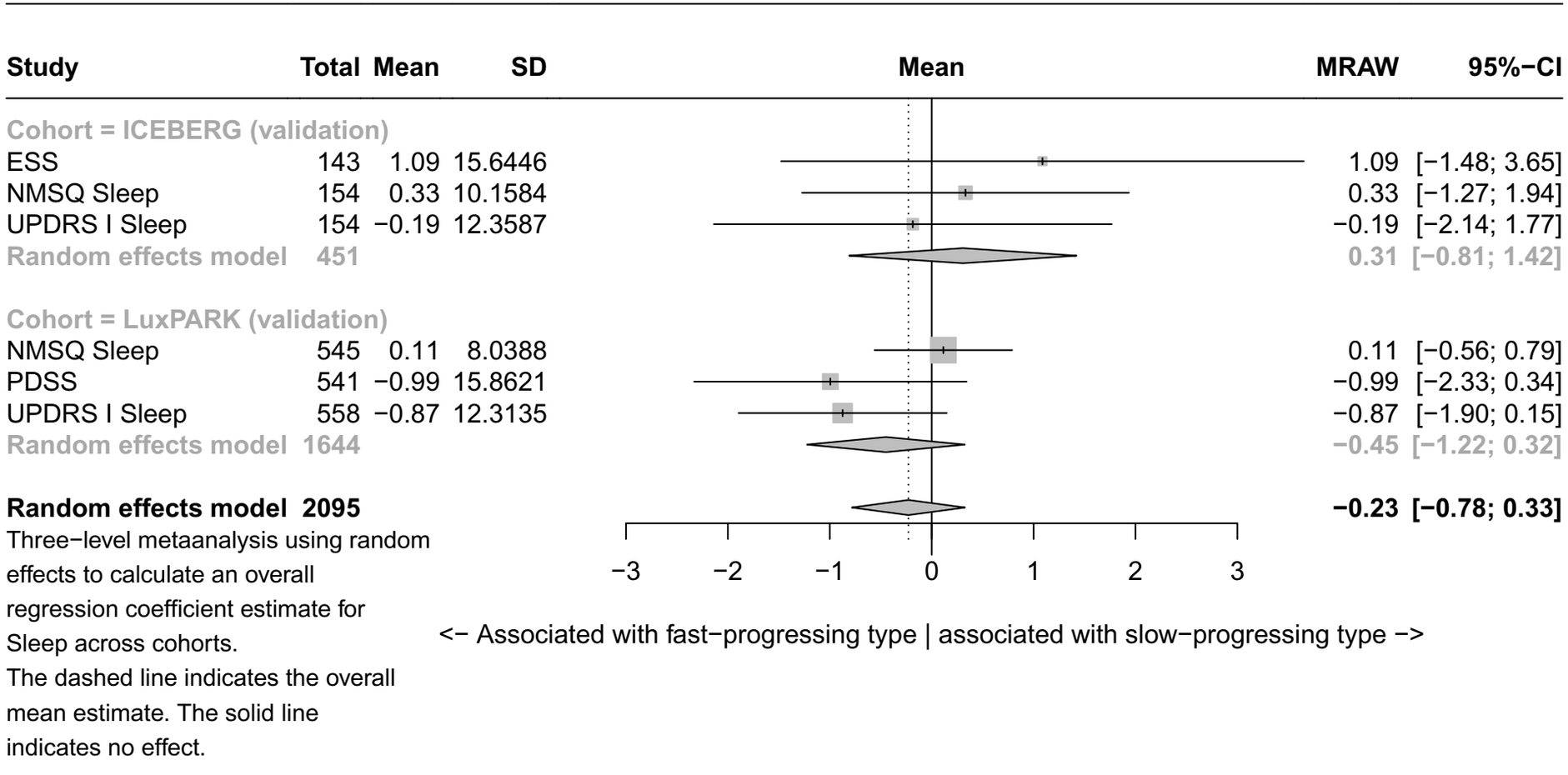
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Overall severity across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Apathy (validation)



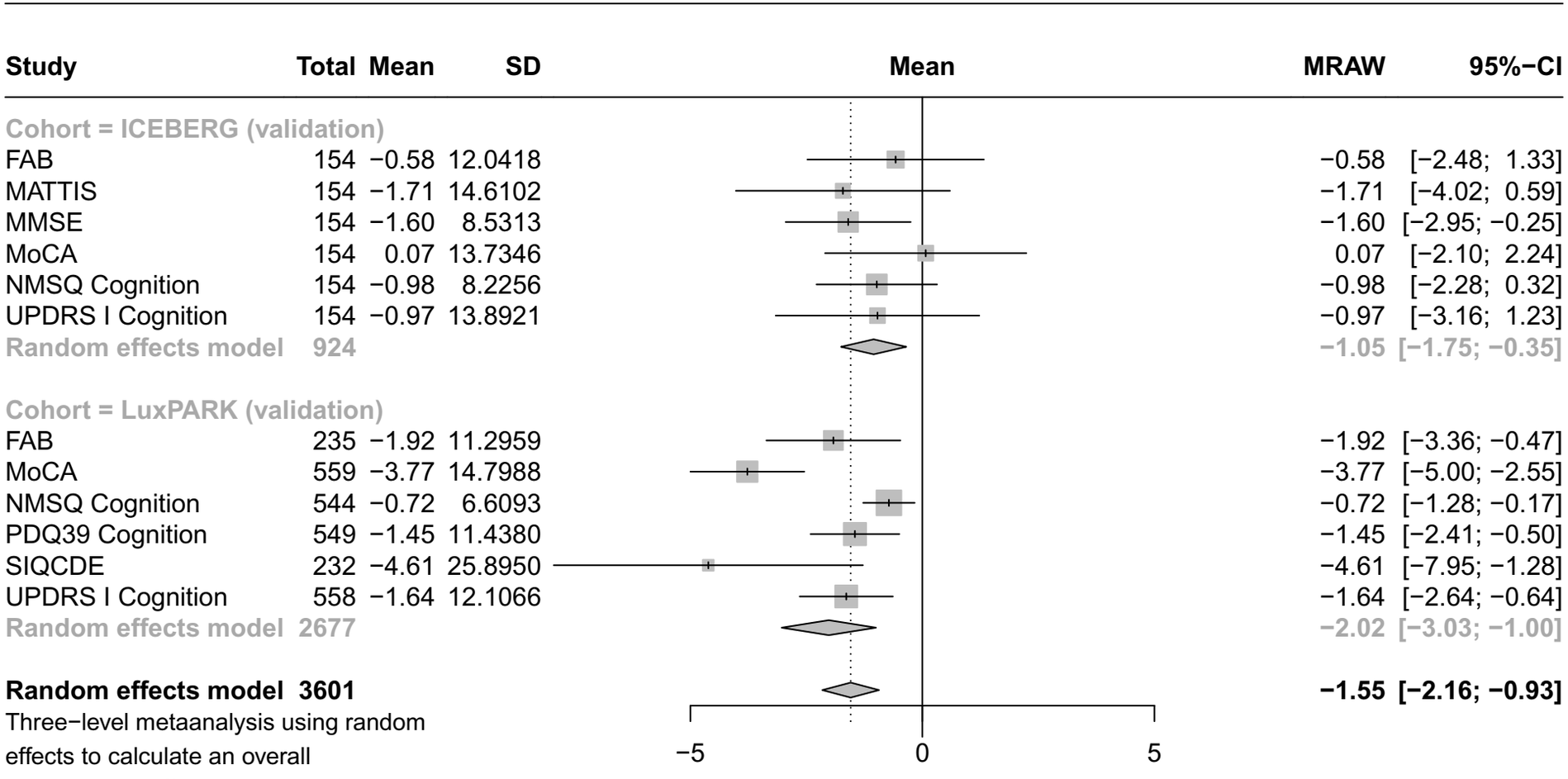
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Apathy across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Sleep (validation)



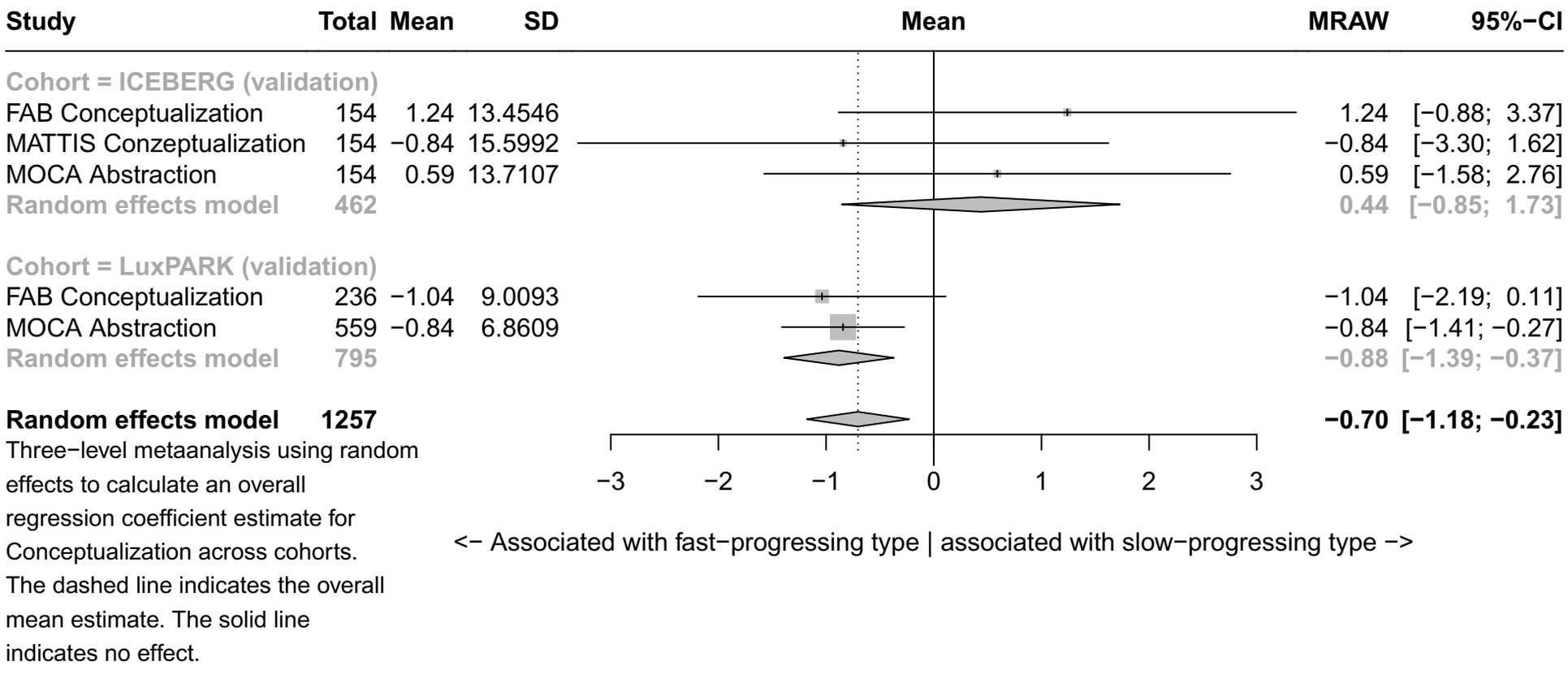
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Sleep across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Overall cognition (validation)



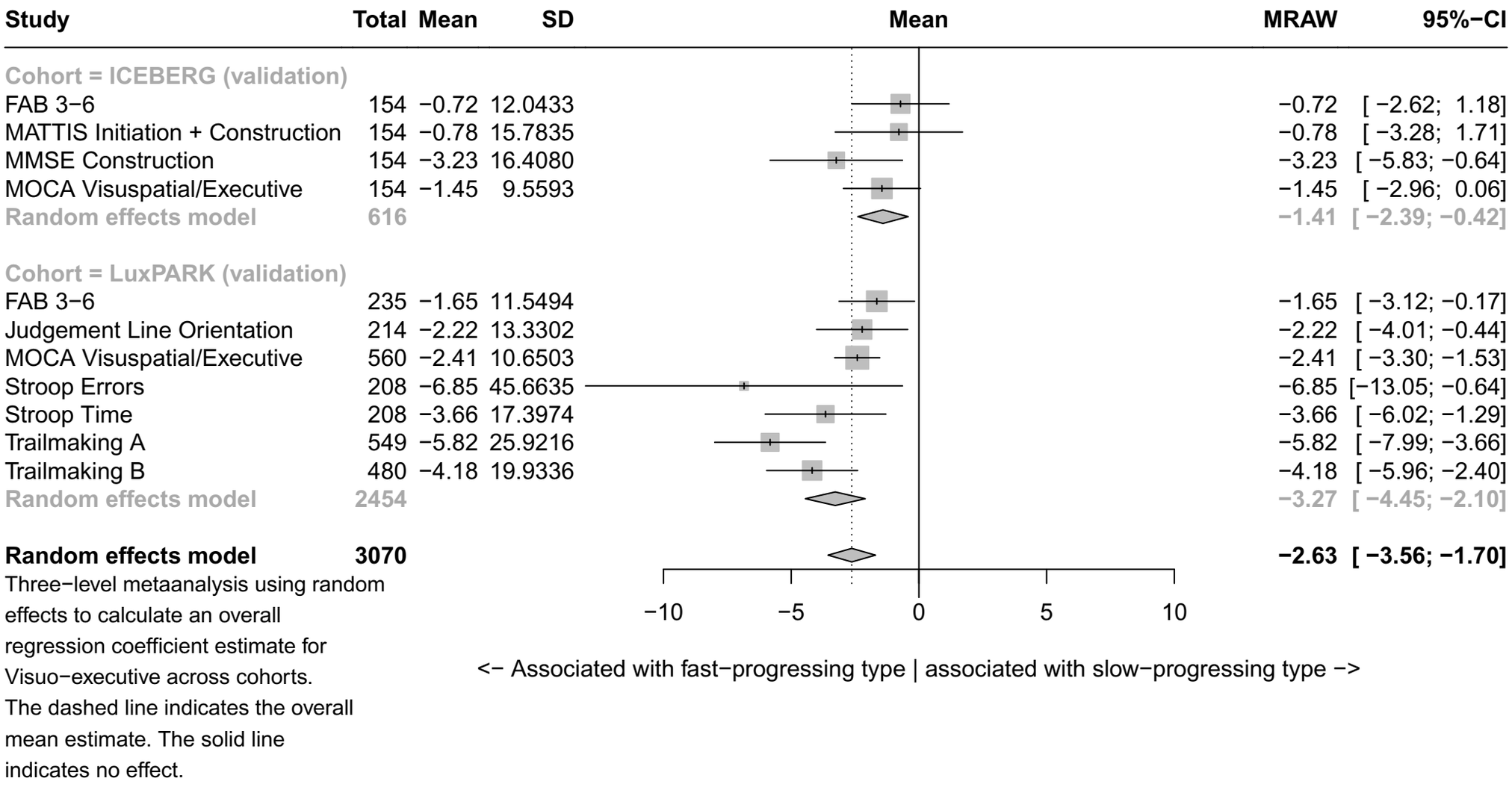
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Overall cognition across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Conceptualization (validation)



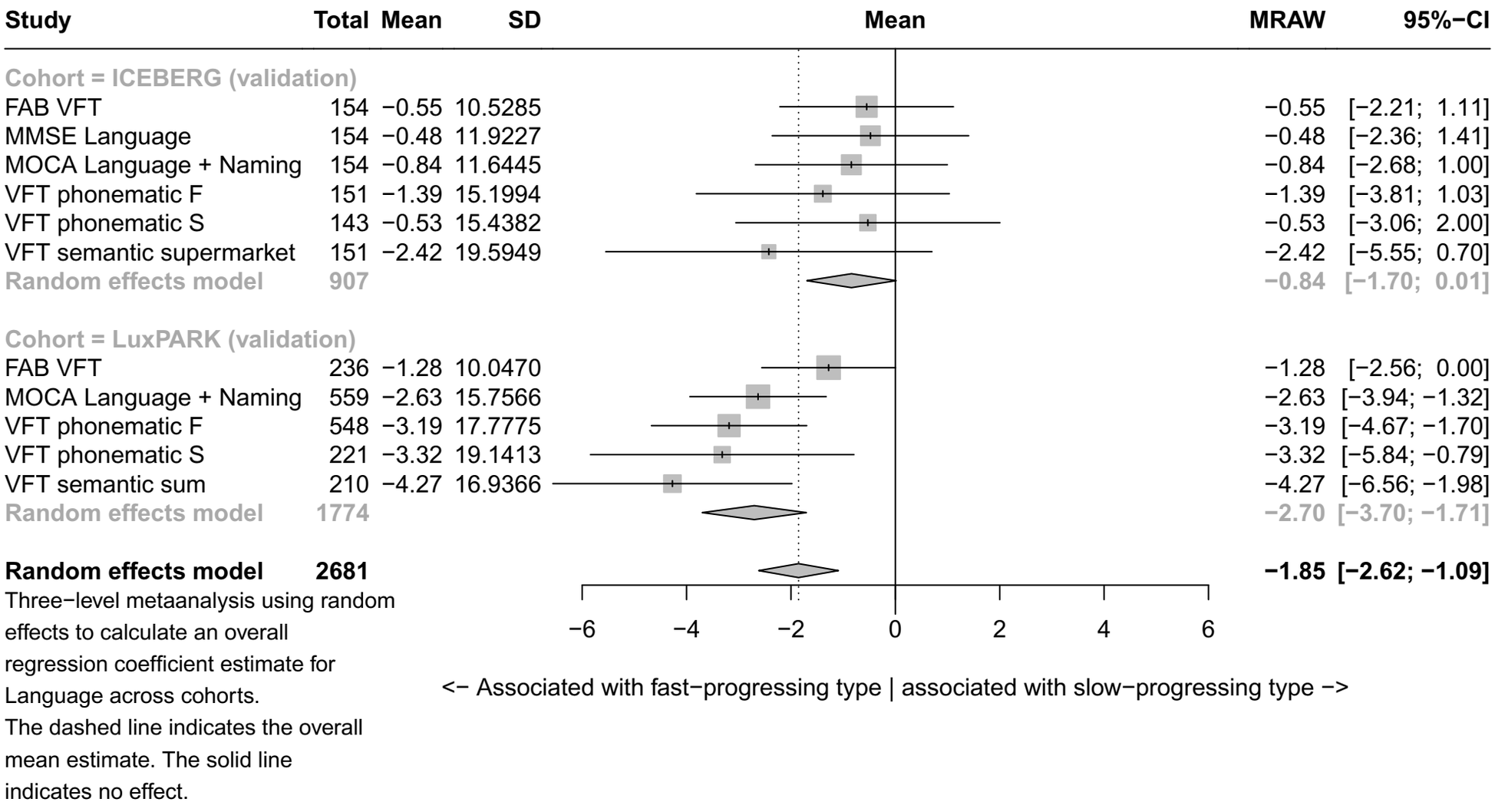
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Conceptualization across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Visuo−executive (validation)



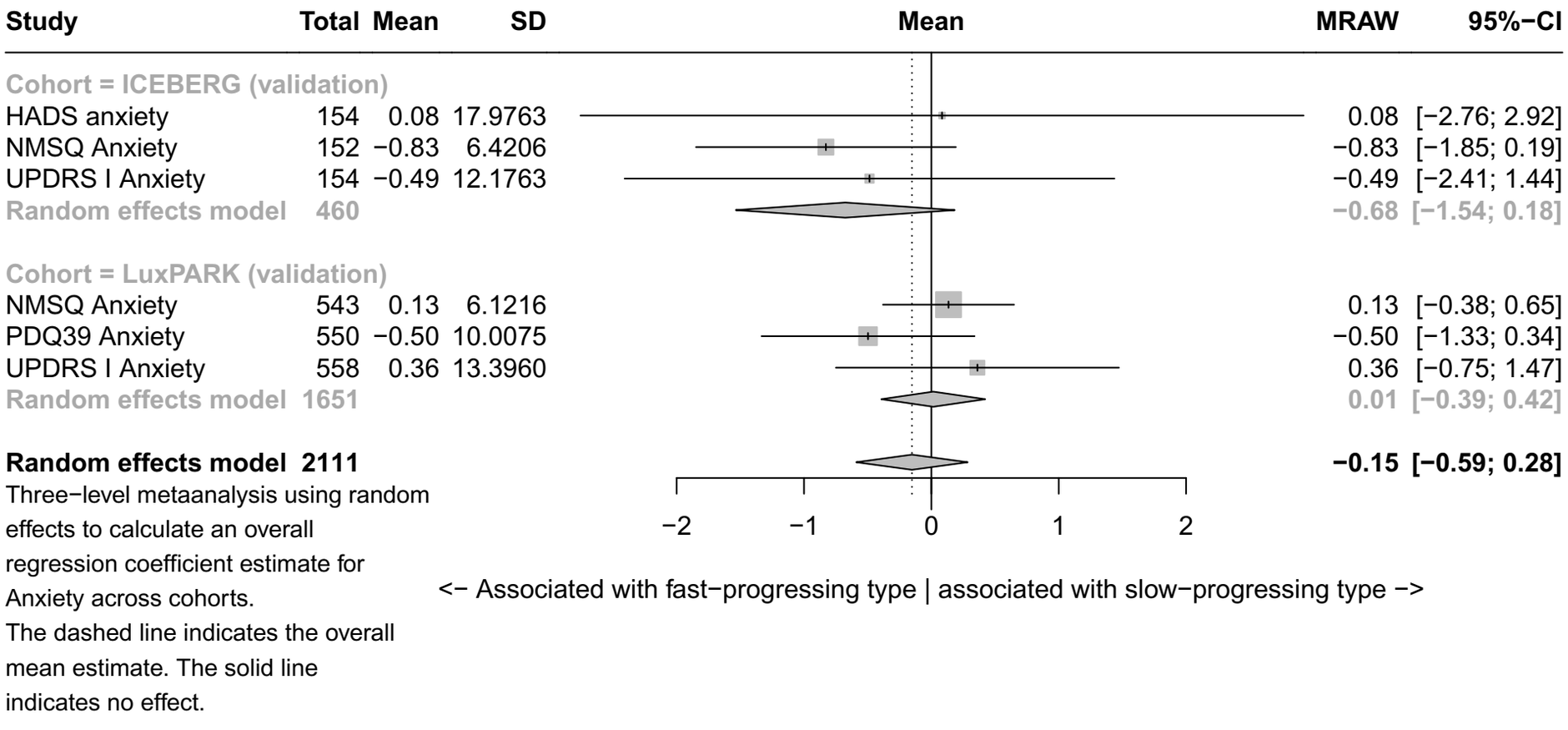
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Visuo−executive across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Language (validation)



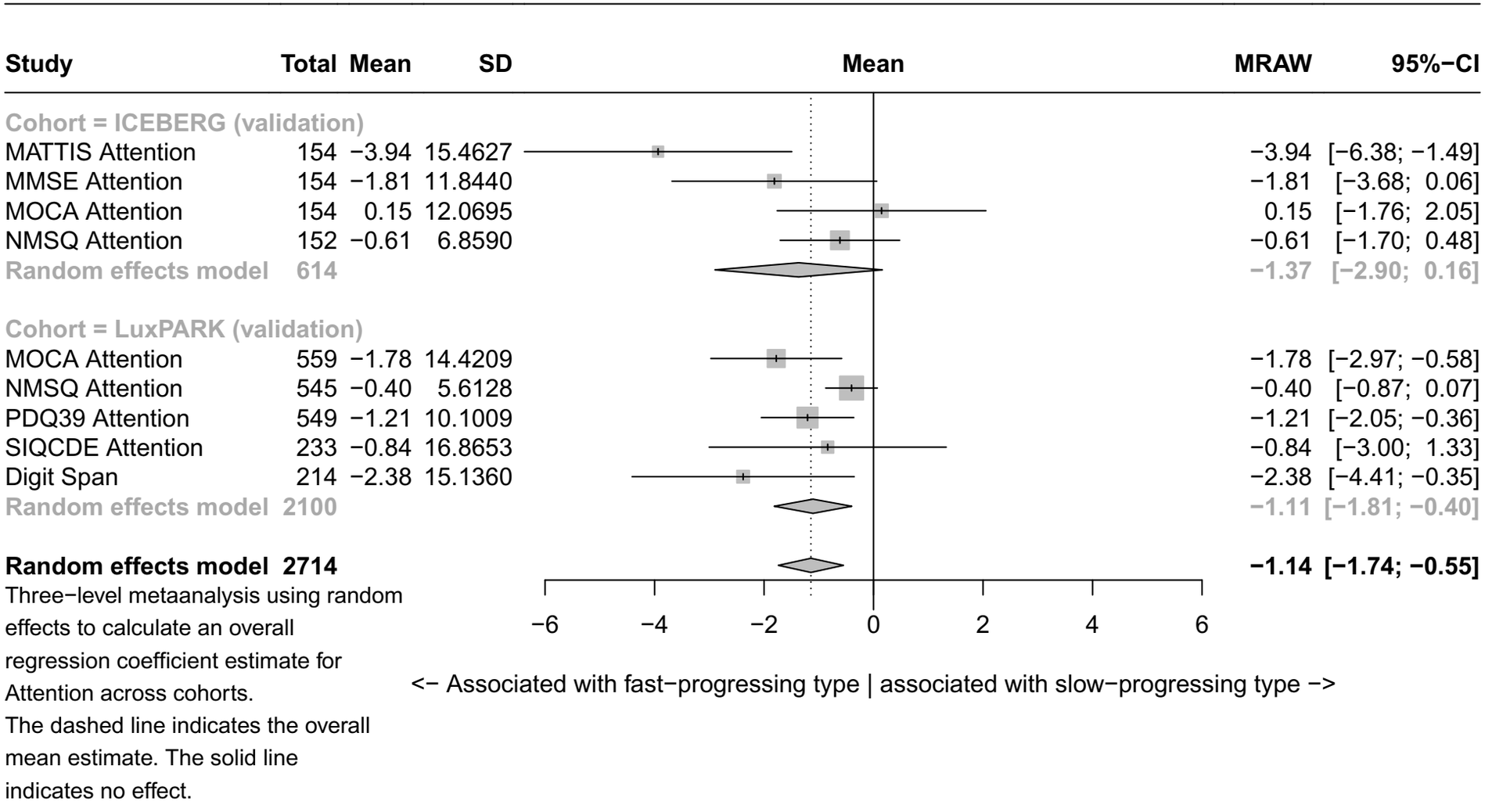
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Language across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Anxiety (validation)



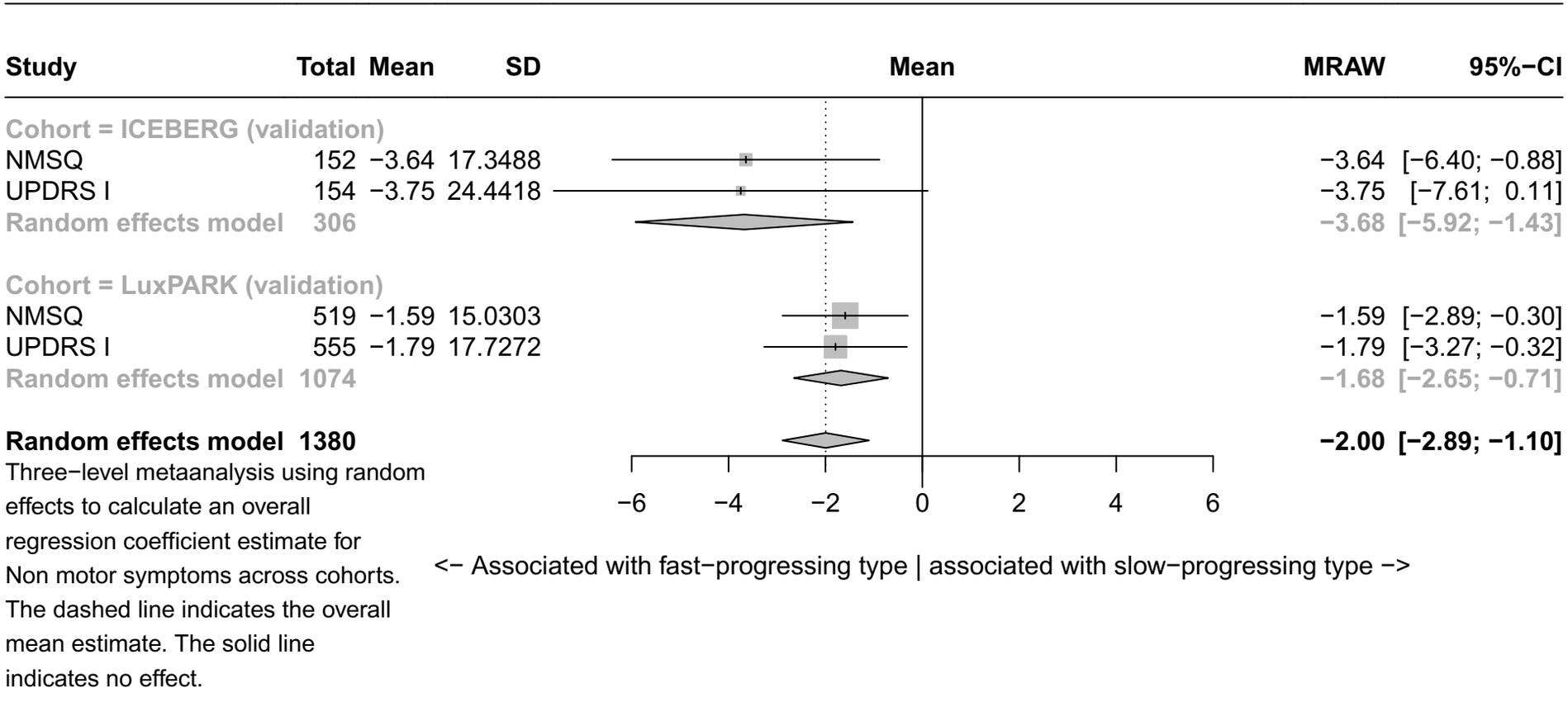
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Anxiety across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Attention (validation)



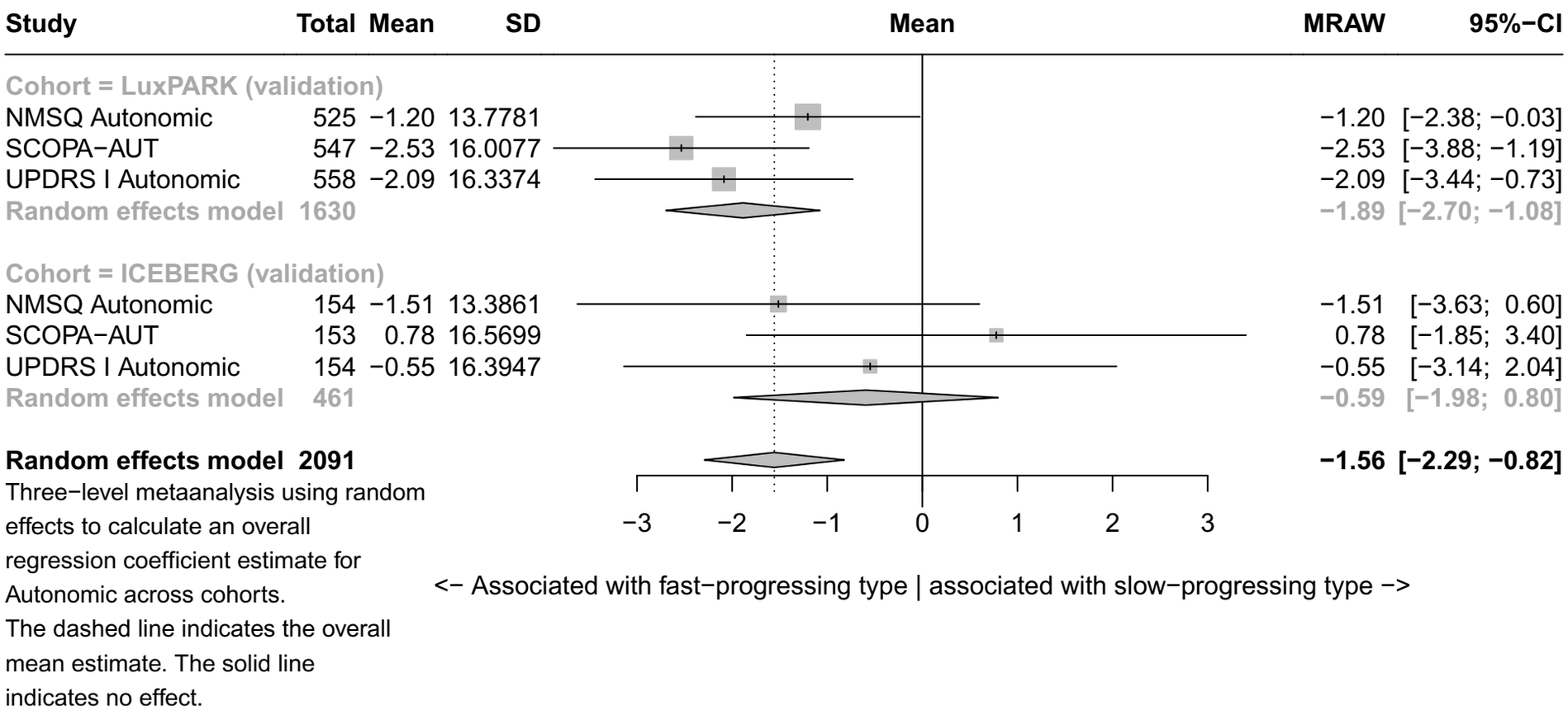
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Attention across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Non motor symptoms (validation)



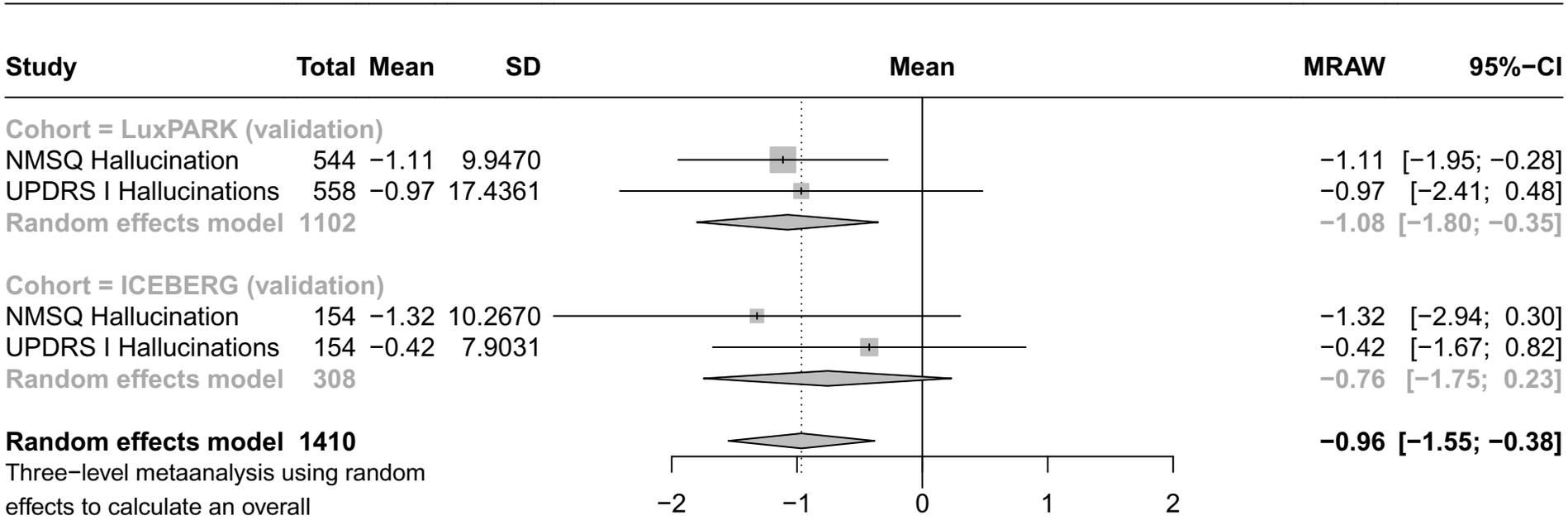
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Non motor symptoms across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Autonomic (validation)



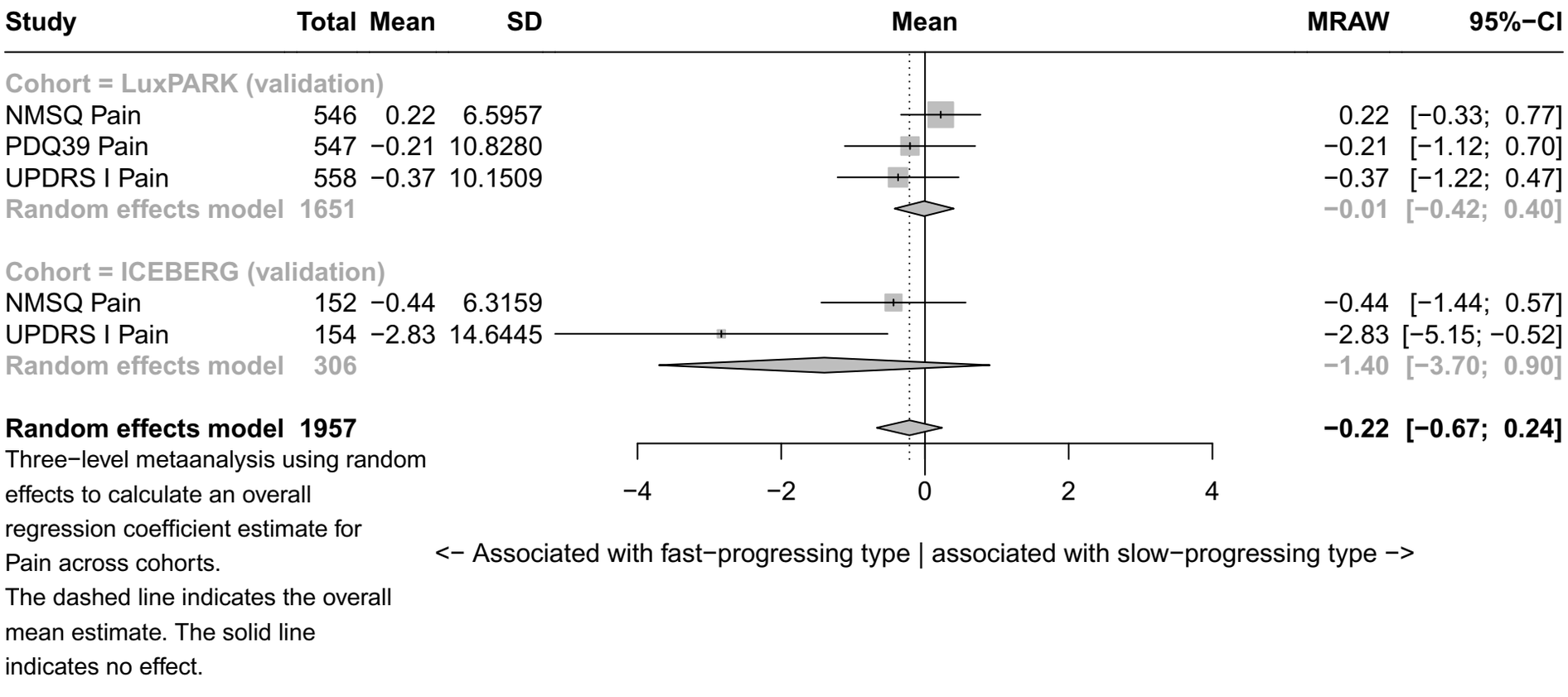
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Autonomic across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Hallucinations (validation)



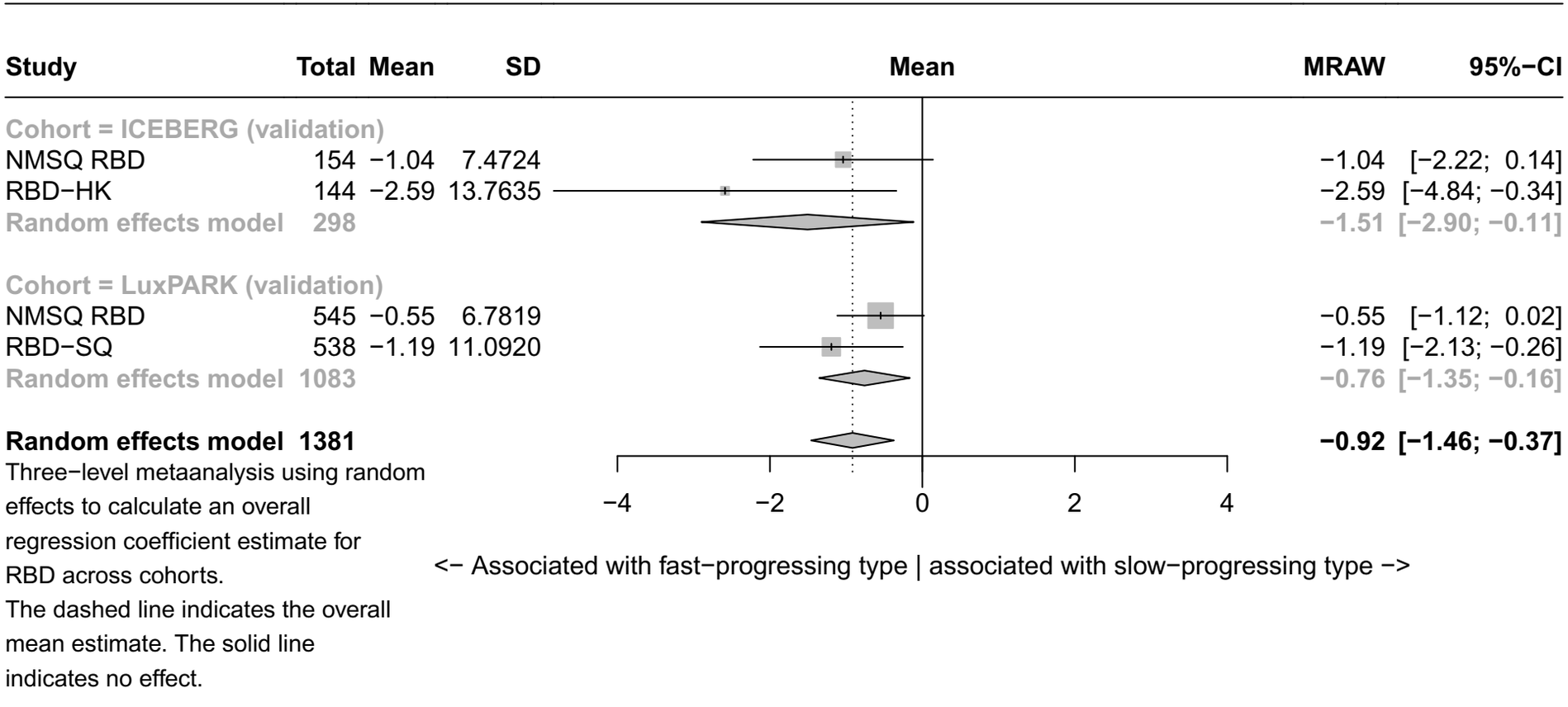
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Hallucinations across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Pain (validation)



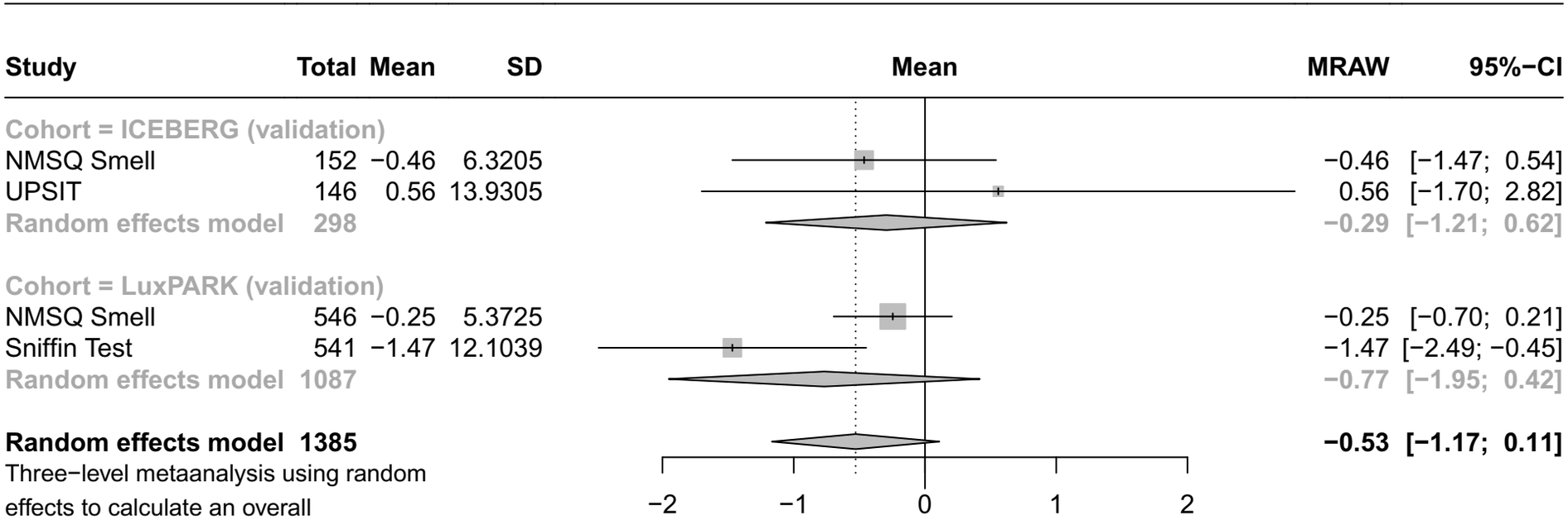
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Pain across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain RBD (validation)



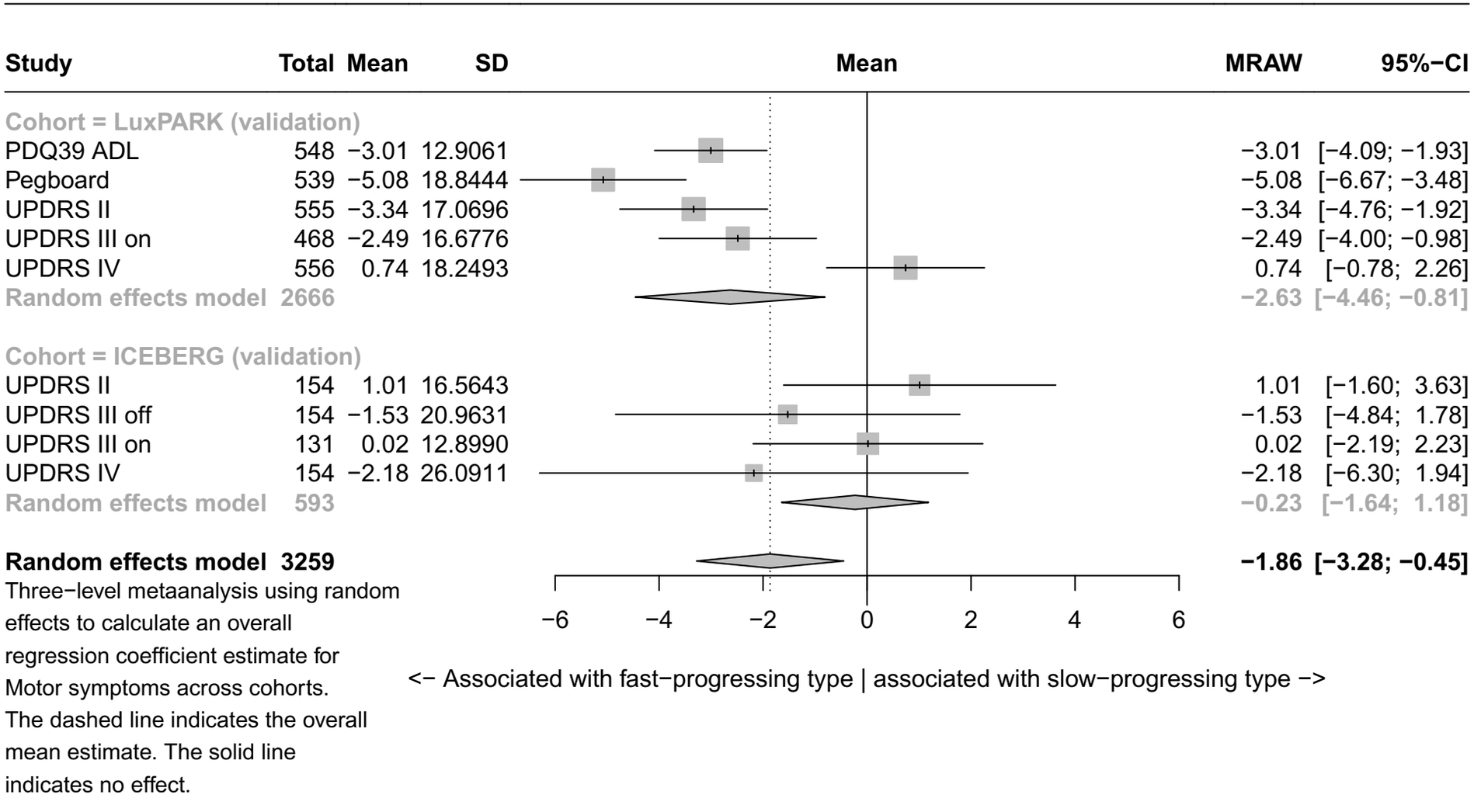
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for RBD across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Smell (validation)



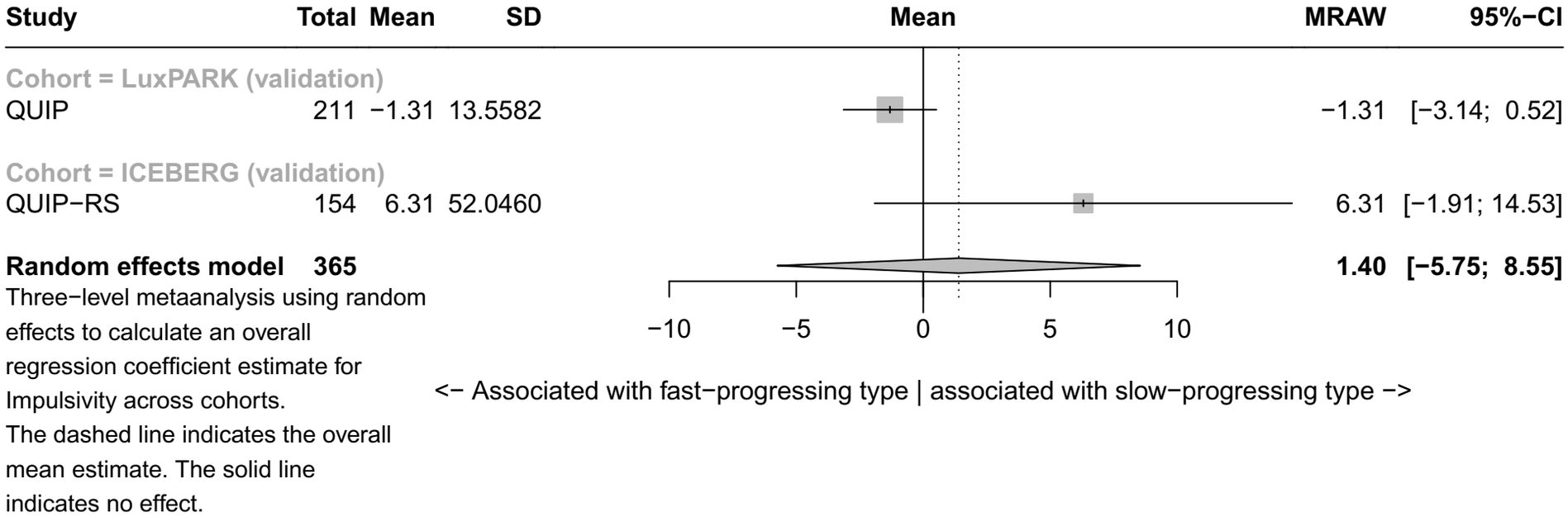
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Smell across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Motor symptoms (validation)



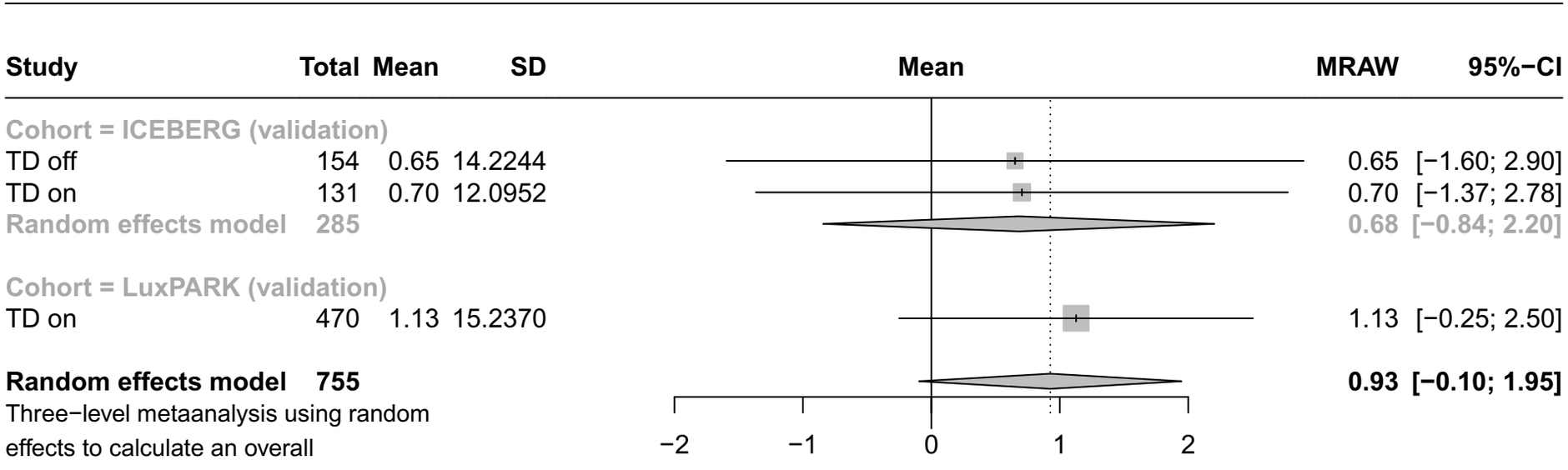
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Motor symptoms across cohorts. The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Impulsivity (validation)



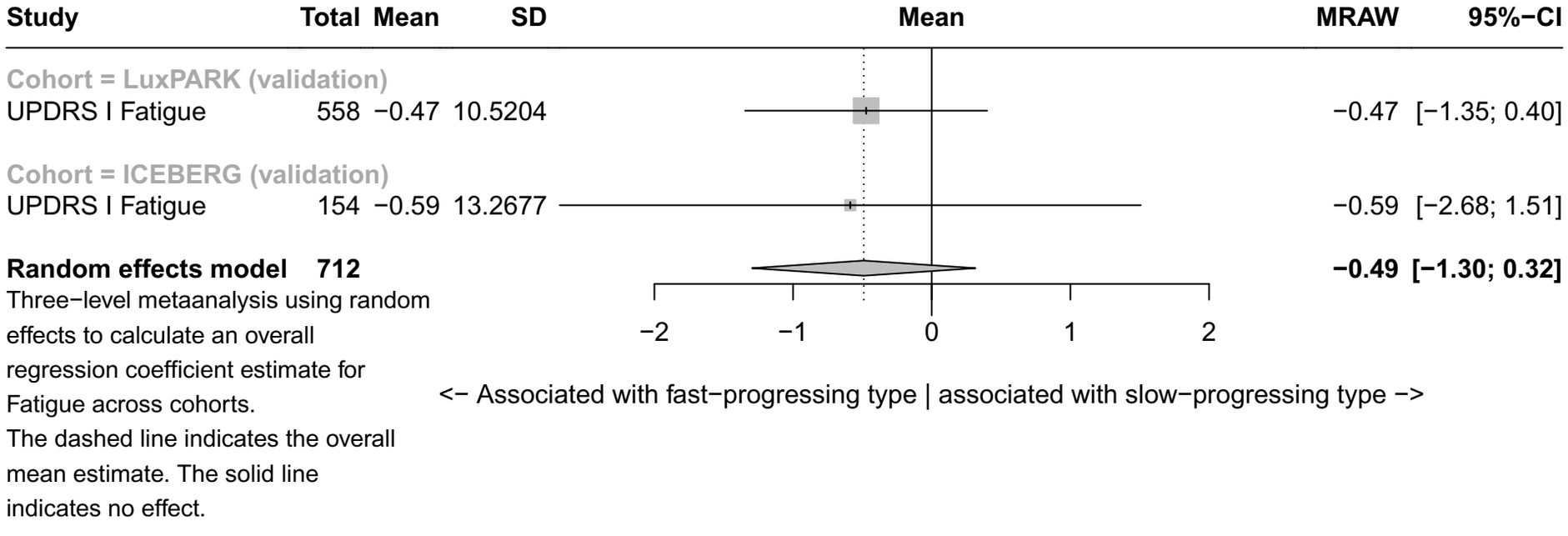
Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Impulsivity across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Tremor (validation)



Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Tremor across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Forest plot for baseline characteristics of symptom domain Fatigue (validation)



Three−level metaanalysis using random effects to calculate an overall regression coefficient estimate for Fatigue across cohorts.
The dashed line indicates the overall mean estimate. The solid line indicates no effect.

Progression subtypes in Parkinson's disease identified by a data driven multi cohort analysis

ICEBERG study group

Steering committee: Marie Vidailhet, MD, PhD, (Pitié-Salpêtrière Hospital, Paris, principal investigator of ICEBERG), Jean-Christophe Corvol, MD, PhD (Pitié-Salpêtrière Hospital, Paris, scientific lead), Isabelle Arnulf, MD, PhD (Pitié-Salpêtrière Hospital, Paris, member of the steering committee), Stéphane Lehericy, MD, PhD (Pitié-Salpêtrière Hospital, Paris, member of the steering committee);

Clinical data: Marie Vidailhet, MD, PhD, (Pitié-Salpêtrière Hospital, Paris, coordination), Graziella Mangone, MD, PhD (Pitié-Salpêtrière Hospital, Paris, co-coordination), Jean-Christophe Corvol, MD, PhD (Pitié-Salpêtrière Hospital, Paris), Isabelle Arnulf, MD, PhD (Pitié-Salpêtrière Hospital, Paris), Sara Sambin, MD (Pitié-Salpêtrière Hospital, Paris), Poornima Menon, MD (Pitié-Salpêtrière Hospital, Paris, Jonas Ihle, MD (Pitié-Salpêtrière Hospital, Paris), Caroline Weill, MD, (Pitié-Salpêtrière Hospital, Paris), David Grabli, MD, PhD (Pitié-Salpêtrière Hospital, Paris); Florence Cormier-Dequaire, MD (Pitié-Salpêtrière Hospital, Paris); Louise Laure Mariani, MD, PhD (Pitié-Salpêtrière Hospital, Paris), Bertrand Degos, MD, PhD (Avicenne Hospital, Bobigny);

Neuropsychological data: Richard Levy, MD (Pitié-Salpêtrière Hospital, Paris, coordination), Fanny Pineau, MS (Pitié-Salpêtrière Hospital, Paris, neuropsychologist), Julie Socha, MS (Pitié-Salpêtrière Hospital, Paris, neuropsychologist), Eve Benchetrit, MS (La Timone Hospital, Marseille, neuropsychologist), Virginie Czernecki, MS (Pitié-Salpêtrière Hospital, Paris, neuropsychologist), Marie-Alexandrine, MS (Pitié-Salpêtrière Hospital, Paris, neuropsychologist);

Eye movement: Sophie Rivaud-Pechoux, PhD (ICM, Paris, coordination); Elodie Hainque, MD, PhD (Pitié-Salpêtrière Hospital, Paris);

Sleep assessment: Isabelle Arnulf, MD, PhD (Pitié-Salpêtrière Hospital, Paris, coordination), Smaranda Leu Semenescu, MD (Pitié-Salpêtrière Hospital, Paris), Pauline Dodet, MD (Pitié-Salpêtrière Hospital, Paris);

Genetic data: Jean-Christophe Corvol, MD, PhD (Pitié-Salpêtrière Hospital, Paris, coordination), Graziella Mangone, MD, PhD (Pitié-Salpêtrière Hospital, Paris, co-coordination), Samir Bekadar, MS (Pitié-Salpêtrière Hospital, Paris, biostatistician), Alexis Brice, MD (ICM, Pitié-Salpêtrière Hospital, Paris), Suzanne Lesage, PhD (INSERM, ICM, Paris, genetic analyses);

Metabolomics: Fanny Mochel, MD, PhD (Pitié-Salpêtrière Hospital, Paris, coordination), Farid Ichou, PhD (ICAN, Pitié-Salpêtrière Hospital, Paris), Vincent Perlbarg, PhD, Pierre and Marie Curie University), Benoit Colsch, PhD (CEA, Saclay), Arthur Tenenhaus, PhD (Supelec, Gif-sur-Yvette, data integration);

Brain MRI data: Stéphane Lehericy, MD, PhD (Pitié-Salpêtrière Hospital, Paris, coordination), Rahul Gaurav, MS, (Pitié-Salpêtrière Hospital, Paris, data analysis), Nadya Pyatigorskaya, MD, PhD, (Pitié-Salpêtrière Hospital, Paris, data analysis); Lydia Yahia-Cherif, PhD (ICM, Paris, Biostatistics), Romain Valabregue, PhD (ICM, Paris, data analysis), Cécile Galléa, PhD (ICM, Paris);

Datscan imaging data: Marie-Odile Habert, MCU-PH (Pitié-Salpêtrière Hospital, Paris, coordination);

Voice recording: Dijana Petrovska, PhD (Telecom Sud Paris, Evry, coordination), Laetitia Jeancolas, MS (Telecom Sud Paris, Evry);

Study management: Alizé Chalançon (Pitié-Salpêtrière Hospital, Paris, Project manager), Carole Dongmo-Kenfack (Pitié-Salpêtrière Hospital, Paris, clinical research assistant); Christelle Laganot (Pitié-Salpêtrière Hospital, Paris, clinical research assistant), Valentine Maheo (Pitié-Salpêtrière Hospital, Paris, clinical research assistant), Manon Gomes (Pitié-Salpêtrière Hospital, Paris, clinical research assistant)

Study sponsoring: The ICEBERG Study was funded by the Programme d'investissements d'avenir (ANR-10-IAIHU-06), the Paris Institute of Neurosciences – IHU (IAIHU-06), the Agence Nationale de la Recherche (ANR-11-INBS-0006), and Électricité de France (Fondation d'Entreprise EDF).

Progression subtypes in Parkinson's disease identified by a data driven multi cohort analysis

NCER-PD/LuxPARK consortium

We would like to thank all participants of the Luxembourg Parkinson's Study for their important support to our research. Furthermore, we acknowledge the joint effort of the National Centre of Excellence in Research on Parkinson's Disease (NCER-PD) Consortium members from the partner institutions Luxembourg Centre for Systems Biomedicine, Luxembourg Institute of Health, Centre Hospitalier de Luxembourg, and Laboratoire National de Santé generally contributing to the Luxembourg Parkinson's Study as listed below:

Geeta ACHARYA[2], Gloria AGUAYO[2], Myriam ALEXANDRE[2], Muhammad ALI[1], Wim AMMERLANN[2], Giuseppe ARENA[1], Rudi BALLING[1], Michele BASSIS[1], Katy BEAUMONT[2], Regina BECKER[1], Camille BELLORA[2], Guy BERCHEM[3], Daniela BERG[11], Alexandre BISDORFF[5], Ibrahim BOUSSAAD[1], Kathrin BROCKMANN[11], Jessica CALMES[2], Lorieza CASTILLO[2], Gessica CONTESOTTO[2], Nico DIEDERICH[3], Rene DONDELINGER[5], Daniela ESTEVES[2], Guy FAGHERAZZI[2], Jean-Yves FERRAND[2], Manon GANTENBEIN[2], Thomas GASSER[11], Piotr GAWRON[1], Soumyabrata GHOSH[1], Marijus GIRAITIS[2,3], Enrico GLAAB[1], Elisa GÓMEZ DE LOPE[1], Jérôme GRAAS[2], Mariella GRAZIANO[17], Valentin GROUES[1], Anne GRÜNEWALD[1], Wei GU[1], Gaël HAMMOT[2], Anne-Marie HANFF[2,20,21], Linda HANSEN[1,3], Michael HENEKA[1], Estelle HENRY[2], Sylvia HERBRINK[6], Sascha HERZINGER[1], Michael HEYMANN[2], Michele HU[8], Alexander HUNDT[2], Nadine JACOBY[18], Jacek JAROSLAW LEBIODA[1], Yohan JAROSZ[1], Sonja JÓNSDÓTTIR[2], Quentin KLOPFENSTEIN[1], Jochen KLUCKEN[1,2,3], Rejko KRÜGER[1,2,3], Pauline LAMBERT[2], Zied LANDOULSI[1], Roseline LENTZ[7], Inga LIEPELT[11], Robert LISZKA[14], Laura LONGHINO[3], Victoria LORENTZ[2], Paula Cristina LUPU[2], Tainá M. MARQUES[1], Clare MACKAY[10], Walter MAETZLER[15], Katrin MARCUS[13], Guilherme MARQUES[2], Patricia MARTINS CONDE[1], Patrick MAY[1], Deborah MCINTYRE[2], Chouaib MEDIOUNI[2], Francoise MEISCH[1], Myriam MENSTER[2], Maura MINELLI[2], Michel MITTELBRONN[1,4], Brit MOLLENHAUER[12], Friedrich MÜHLSCHLEGEL[4], Romain NATI[3], Ulf NEHRBASS[2], Sarah NICKELS[1], Beatrice NICOLAI[3], Jean-Paul NICOLAY[19], Fozia NOOR[2], Marek OSTASZEWSKI[1], Clarissa P. C. GOMES[1], Sinthuja PACHCHEK[1], Claire PAULY[1,3], Laure PAULY[2,20], Lukas PAVELKA[1,3], Magali PERQUIN[2], Nancy E. RAMIA[1,] Rosalina RAMOS LIMA[2], Armin RAUSCHENBERGER[1], Rajesh RAWAL[1], Dheeraj REDDY BOBBILI[1], Kirsten ROOMP[1], Eduardo ROSALES[2], Isabel ROSETY[1], Estelle SANDT[2], Stefano SAPIENZA[1], Venkata SATAGOPAM[1], Margaux SCHMITT[2], Sabine SCHMITZ[1], Reinhard SCHNEIDER[1], Jens SCHWAMBORN[1], Amir SHARIFY[2], Ekaterina SOBOLEVA[1], Kate SOKOLOWSKA[2], Hermann THIEN[2], Elodie THIRY[3], Rebecca TING JIIN LOO[1], Christophe TREFOIS[1], Johanna TROUET[2], Olena TSURKALENKO[2], Michel VAILLANT[2], Mesele VALENTI[2], Gilles VAN CUTSEM[1,3], Carlos VEGA[1], Liliana VILAS BOAS[3], Maharshi VYAS[1], Richard WADE-MARTINS[9], Paul WILMES[1], Evi WOLLSCHEID-LENGELING[1], Gelani ZELIMKHANOV[3]

1. Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
2. Luxembourg Institute of Health, Strassen, Luxembourg
3. Centre Hospitalier de Luxembourg, Strassen, Luxembourg
4. Laboratoire National de Santé, Dudelange, Luxembourg
5. Centre Hospitalier Emile Mayrisch, Esch-sur-Alzette, Luxembourg
6. Centre Hospitalier du Nord, Ettelbrück, Luxembourg
7. Parkinson Luxembourg Association, Leudelange, Luxembourg
8. Oxford Parkinson's Disease Centre, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
9. Oxford Parkinson's Disease Centre, Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road, Oxford, UK
10. Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, UK
11. Center of Neurology and Hertie Institute for Clinical Brain Research, Department of Neurodegenerative Diseases, University Hospital Tübingen, Germany
12. Paracelsus-Elena-Klinik, Kassel, Germany
13. Ruhr-University of Bochum, Bochum, Germany
14. Westpfalz-Klinikum GmbH, Kaiserslautern, Germany
15. Department of Neurology, University Medical Center Schleswig-Holstein, Kiel, Germany

16. Department of Neurology Philipps, University Marburg, Marburg, Germany
17. Association of Physiotherapists in Parkinson's Disease Europe, Esch-sur-Alzette, Luxembourg
18. Private practice, Ettelbruck, Luxembourg
19. Private practice, Luxembourg, Luxembourg
20. Faculty of Science, Technology and Medicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
21. Department of Epidemiology, CAPHRI School for Public Health and Primary Care, Maastricht University Medical Centre+, Maastricht, the Netherlands