<u>Supplement 1</u>

Adverse Childhood Experiences

1. poverty
2. parental arrests
3. parental hospitalizations
4. accident or crash with automobile, plane, or boat
5. accidental burning, poisoning, or drowning
6. attacked by an animal
7. adult loved one died
8. sibling or peer died
9. domestic violence
10. hospitalized, visited ER, or invasive medical procedure
11. man-made disaster
12. natural disaster
13. other life event
14. physical abuse
15. sexual abuse, sexual assault, or rape
16. witnessed someone threatened with harm, seriously injured, or killed

PAPA/CAPA Irritability Item Text

PGE0I01 LOSING TEMPER
Discrete episodes of temper manifested by shouting or name calling but without violence and not meeting criteria for a temper tantrum.

- *What sort of temper has s/he got?*
- *What happens when s/he loses his/her temper?*

0 = Absent
2 = Present

PGE1I01 NON-DESTRUCTIVE TEMPER TANTRUMS
Discrete episodes of excessive temper, frustration or upset, manifested by shouting, crying or stamping or non- destructive violence directed against property.
Violence or damage done here does not constitute Vandalism or Assault.

- *What sort of temper has s/he had in the past 3 months?*
- *What happens when something upsets him/her or s/he doesn't get what s/he wants?*
- *Does s/he have angry outbursts?*
- *Does s/he have temper tantrums?*

0 = Absent
2 = Excessive temper, upset, shouting, crying or non destructive violence directed only against property, (e.g. stamping, kicking, throwing toys, hitting walls, spitting, holding breath, etc.).

PDA8I01 IRRITABILITY
Increased ease of precipitation of externally directed feelings of anger, bad temper, short temper, resentment, or annoyance. (Change may predate the primary period and continue into at least part of the primary period.)
Note that this rating is of a change in the child's usual liability to be precipitated into anger, it does not refer to the form of the anger once it has been precipitated.
N.B.: The irritable mood itself is being rated, not just its manifestations; thus, frequency and duration ratings refer to the number and length of episodes of the mood, not of the episodes of snappiness, shouting or quarrelsomeness.

- *Has s/he been more irritable than usual in the last 3 months?*
- *Or made angry more easily?*
- *Has s/he had more tantrums than usual in the last 3 months?*

0 = Absent
2 = Irritable mood present in at least 2 activities manifested by at least one instance of snappiness, shouting, quarrelsomeness and at least sometimes uncontrollable by child.
3 = Irritable mood present in most activities, accompanied by snappiness, shouting, quarrelsomeness, and nearly always uncontrollable by child.

PDA6I01 TOUCHY OR EASILY ANNOYED

The child is generally more prone to FEELINGS of anger bad temper, short temper, resentment, sulking or annoyance, UNDER MINOR PROVOCATION than most children. This pattern need not represent a change in behavior.

- *Do things get on his/her nerves easily?*
- *What sorts of things?*
- *Does s/he get annoyed more easily than most children, do you think?*

0 = Absent
2 = Present

PDA7I01 ANGRY OR RESENTFUL
The child is generally more prone to MANIFESTATIONS of anger or resentment (such as snappiness, shouting, quarreling or sulking) under minor provocation, than most children. This pattern need not represent a change in behavior.

- *Does s/he get angry very often?*
- *Does s/he get "sulky" or "pout"?*

0 = Absent
2 = Present

Examining Trajectories of Irritability Across Development

In the latent class mixture models, irritability was the dependent variable and age at each assessment (in months) was the independent variable. Age was indicated to have a random and class-specific effect. A linear link function was used and the variance-covariance matrix for the random effect was diagonal.

An automatic grid search was used to run the estimation function for a maximum of 20 iterations from 50 random sets of initial values to avoid convergence on local maxima. At the end, the parameters corresponding to the best log-likelihood were used as initial values for the final estimation of the parameters.

MRI Acquisition and Structural Analysis

Three waves of neuroimaging data were collected using a 3T TIM TRIO Siemens scanner. At each wave, two T1-weighted images were acquired in the sagittal plane using an MPRAGE 3D sequence (TR=2400ms, TE=3.16ms, flip angle=8°, slab=176mm, 176 slices, matrix size=256×256, field of view=256mm, voxel size=1×1×1mm). The T1 image with the best image quality was used for structural analyses.

Structural images were processed using the longitudinal stream in FreeSurfer, version 5.3 (http://surfer.nmr.mgh.harvard.edu/ [1]). Processing included skull stripping, Talairach transformations, atlas registration, and creating spherical surface maps and parcellations, initialized with common information from an unbiased within-patient template. This longitudinal stream reduces biases that could occur by selecting a single image as a baseline for registration and increases reliability and statistical power significantly [2]. White and pial surfaces generated by FreeSurfer were visually inspected by a single trained and supervised staff member. Surface edits were performed as needed and the surfaces were regenerated as is recommended by FreeSurfer for quality control. From the overall PDS imaging sample, approximately 10% of scans had to be discarded for poor scan quality. These data were resampled on to an average template (fsaverage) and smoothed with a 10mm full-width/half-max kernel.

1. Reuter M, Schmansky NJ, Rosas HD, Fischl B. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage.* 2012;61(4):1402-1418.
2. Reuter M, Fischl B. Avoiding asymmetry-induced bias in longitudinal image processing. *Neuroimage.* 2011;57(1):19-21.

Testing Interactions with Quadratic Effects of Age

As thickness may change non-linearly with age, we examined a linear mixed model approach in Freesurfer [1, 2] in addition to the analyses noted in the main text. In this model, we included a random intercept for each participant, trajectory group as a between-subject factor and interactions with linear and quadratic age trends across wave. Quadratic interactions between age and trajectory group were examined to test for potential non-linear differences in thickness. No regions passed multiple comparisons correction when examining class x quadratic age trend interactions in these linear mixed models.

1. Bernal-Rusiel JL, Reuter M, Greve DN, Fischl B, Sabuncu MR, Alzheimer's Disease Neuroimaging I (2013): Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data. *Neuroimage*. 81:358-370.
2. Bernal-Rusiel JL, Greve DN, Reuter M, Fischl B, Sabuncu MR, Alzheimer's Disease Neuroimaging I (2013): Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models. *Neuroimage*. 66:249-260.

	Original PDS Sample	Latent Class Subsample	Imaging Subsample
Class N	306	271	139
N Assessments (m/sd)	5.5 (2.0)	6.0 (1.4) ***	6.9 (0.9) *
Sex (n/% Female)	148 (48.4%)	132 (48.7 %)	67 (48.2%)
Race (n/% white)	164 (53.6%)	149 (55.0%)	70 (50.4%)
Age at Baseline (m/sd months)	53.4 (9.6)	53.5 (9.5)	54.4 (9.3)
Age at Scan 1 (m/sd months)	122.8 (14.8)	122.8 (14.8)	123.3 (14.8)
IQ (m/sd)	104.7 (14.7)	104.8 (14.8)	104.5 (15.2)
Baseline Adversity Score (m/sd)	2.7 (1.9)	2.7 (1.9)	2.9 (2.0)
Baseline Income-to-Needs	2.2 (1.3)	2.2 (1.3) *	2.0 (2.1) *
HBQ Peer Relations (m/sd)	3.4 (0.6)	3.4 (0.6)	3.4 (0.6)
HBQ Academic Functioning (m/sd)	3.9 (0.7)	3.9 (0.7)	3.9 (0.7)
HBQ Global Functioning (m/sd)	4.3 (5.1)	4.3 (5.1)	4.4 (5.2)
Maternal MDD (n/%)	119 (38.9%)	108 (39.9%)	65 (46.8%) *
Child MDD (n/%)	152 (49.7%)	138 (50.9%)	82 (88.60%) **
Child Anxiety (n/%)	161 (52.6%)	94 (34.7%) *	53 (38.1%)
Child ADHD (n/%)	107 (35.0%)	152 (56.1%)	82 (59.0%)
Child ODD/CD (n/%)	133 (43.5%)	117 (43.2%)	63 (45.3%)

Table S1. Summarizes the demographic and clinical characteristics of the full Preschool Depression Study (PDS) sample (N=306), the sample of children included in the latent class mixture models (N=271 children with at least 3 waves of PAPA data available), and the subset of those children included in the imaging analyses (N=139 children with at least 2 waves of imaging data).

Note: Significant differences between those retained and excluded from each subsample are denoted in either the latent class or imaging column (*p<.05, **p<.01, ***p<.001). Specifically, chi-squared and t-tests were used to compare children included in the latent class subsample to those excluded from the original PDS sample, i.e. who did not have sufficient PAPA data. Compared to those excluded from the latent class analyses, the included children had more PAPA assessments (based on their inclusion criterion), tended to have higher income to needs ratios, and were more likely to have anxiety. Compared to those included I the latent class but excluded from the imaging analyses, the final 139 children in the imaging analyses completed more assessments, reported lower income to needs ratios, and exhibited more maternal and child depression diagnoses.

	Peak Coordinates						
Region	**X**	**Y**	**Z**	**Size (mm2)**	**F**	**Predicting**	**Interaction**
Right Superior Frontal Gyrus	9	53	11	466	3.36	Average	Class x Sex
Left Superior Parietal Lobule	-15	-82	27	428	3.15	Average	Class x Sex
Left Rostral Middle Frontal Gyrus	-41	27	31	644	3.45	Rate	Class x Age
Right Posterior Cingulate Gyrus	9	-26	30	544	4.28	Rate	Class x Age
Right Precentral Gyrus	11	-17	68	1013	3.80	Rate	Class x Age
Right Cingulate Isthmus	22	-53	6	552	4.36	Rate	Class x Age x Sex
Right Middle Temporal Gyrus	48	0	-32	1159	4.25	Rate	Class x Age x Sex
Right Postcentral Gyrus	52	-13	17	809	4.14	Rate	Class x Age x Sex

Table S2: Regions showing interactions with trajectory class
Note: These regions showed interactions between trajectory class and/or age and sex (Interaction); the F-statistic of this interaction is presented here as well as the peak coordinates and size of each cluster. These effects were significant either in models examining the temporal average of cortical thickness or rate of change (Predicting). As the trajectory classes split by sex or age were small, we did not explore these interactions as they are likely underpowered.

Classes	Log-likelihood	Parameters	BIC	AIC	Entropy	% class 1	% class 2	% class 3	% class 4	% class 5
1	-2685.99	5	5400.00	5129.09	-	100.00				
2	-2672.84	8	5390.49	5116.79	0.39	69.00	31.00			
3	***-2655.31***	***11***	***5372.25***	***5111.86***	***0.68***	***16.24***	***30.63***	***53.14***		
4	-2651.06	14	5380.56	5118.89	0.65	7.75	29.52	23.62	39.11	
5	-2644.71	17	5384.66	5117.89	0.67	8.86	26.20	38.75	15.13	11.07

Table S3. Summarizes the latent class mixture models with one through five classes that were tested.

Note: The result with three trajectory classes was used for subsequent analyses as it showed the lowest Bayesian Information Criterion (BIC), the lowest Akaike Information Criterion (AIC), and the highest entropy. While the models with four or five classes showed slightly lower log-likelihood values, the BIC and AIC values were higher, the entropy was slightly lower, and both had small trajectory classes containing <10% of the sample.

(A) Predicting Later MDD	b	Wald	p	Odds Ratio
Baseline Irritability	*0.31*	*7.32*	*.01*	*1.37*
Sex (F>M)	-0.10	0.10	.75	0.91
Maternal MDD	0.63	4.08	.04	1.88
Baseline MDD Symptoms	*0.54*	*37.76*	*<.001*	*1.71*
Adverse Childhood Experiences	*0.23*	*7.14*	*.01*	*1.26*

(B) Predicting Later Anxiety	b	Wald	p	Odds Ratio
Baseline Irritability	0.02	0.02	.89	1.02
Sex (F>M)	0.37	1.88	.17	1.45
Maternal Anxiety	0.07	0.03	.86	1.07
Baseline Anxiety Symptoms	*0.44*	*38.61*	*<.001*	*1.55*
Adverse Childhood Experiences	0.03	0.13	.71	1.03

Table S4. Presents the logistic regression confidents (b), Wald statistics, significance (p), and odds ratios for the prediction of later major depressive disorder diagnosis (MDD) in (A) or later anxiety disorder diagnosis (B).

Note: Significant predictors (p<.05) are in italics. Baseline irritability was a significant predictor of later MDD above and beyond sex, maternal MDD, baseline MDD symptoms (subtracting out irritability symptoms from the depression module), and adverse childhood experiences. This effect was not significant when predicting anxiety.

	Left Superior Frontal Gyrus	Left Superior Temporal Gyrus	Right Inferior Parietal Lobule	Left Rostral Middle Frontal Gyrus	Right Superior Frontal Gyrus
Age at Scan 1 (years)	-0.149	-0.143	-.203*	-.308**	-.256**
Sex	-0.768	0.064	-0.758	0.059	0.26
Baseline non-irritability depression symptom severity	0.017	-0.021	0.009	-0.076	-0.099
Baseline income to needs ratio	-0.065	-0.058	0.058	0.009	-0.098
IQ	-0.100	-0.121	0.091	-0.051	-0.116
Baseline adverse life events	-0.044	-0.059	-0.106	-0.033	-0.094
Irritability Severity at Scan 1	0.116	0.049	0.030	-0.052	0.010
Maternal depression history	-0.295	-0.408	-0.46	0.337	0.353
Psychotropic medication use ever	-2.022*	-1.504	-0.221	-0.124	-1.512
ADHD diagnosis ever	-1.649	-0.403	0.837	1.714	0.657
MDD diagnosis ever	1.209	-0.372	0.327	1.652	1.18
Anxiety diagnosis ever	-0.441	0.565	0.674	1.097	0.462
ODD/CD diagnosis ever	-1.695	-1.101	0.093	0.981	0.309

*p<.05; **p<.01; ***p<001

Table S5: Associations between Thickness and Covariates
Note: This table displays the associations between factors that varied by trajectory group or that could influence cortical thickness. T-statistics are presented for effects of binary variables (sex, maternal depression history, medication use, child diagnoses). Pearson's correlation coefficients are presented for the remaining variables.

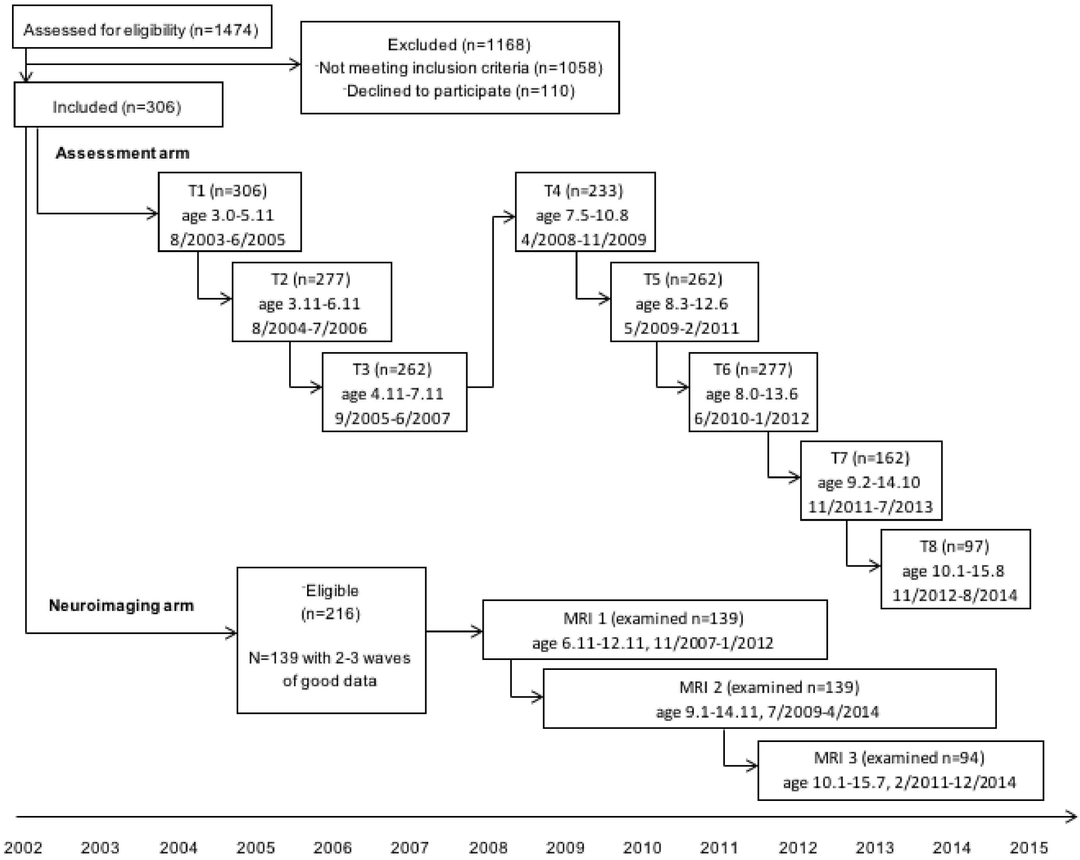

Figure S1: Study Flow Diagram

Note: This figure displays the flow of data through the Preschool Depression Study. The total number of children through each assessment wave is presented. The number of children examined at each wave of imaging is presented here, particularly only children with two or more waves of good quality data were examined.

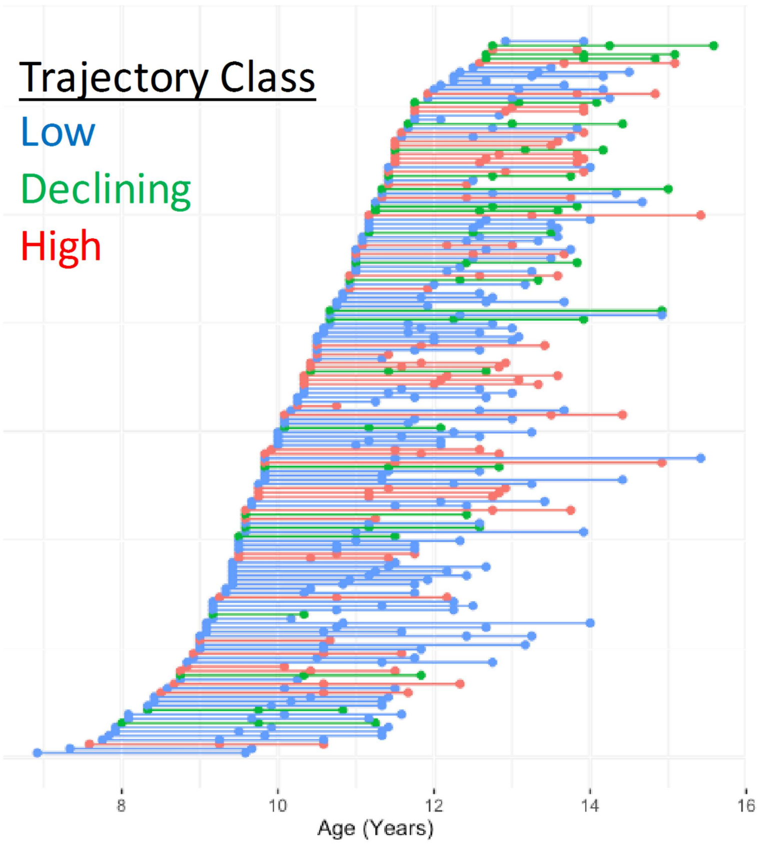

Figure S2. Timing of each scan wave with good structural imaging data is presented here. Note: Points connect each individual by lines. Individuals are sorted by the age at the first scan and colored by irritability trajectory class. The first and second waves were collected on average 18.67 months apart (SD=6.52) and the second and third were similarly spaced (M=15.53, SD=6.14).

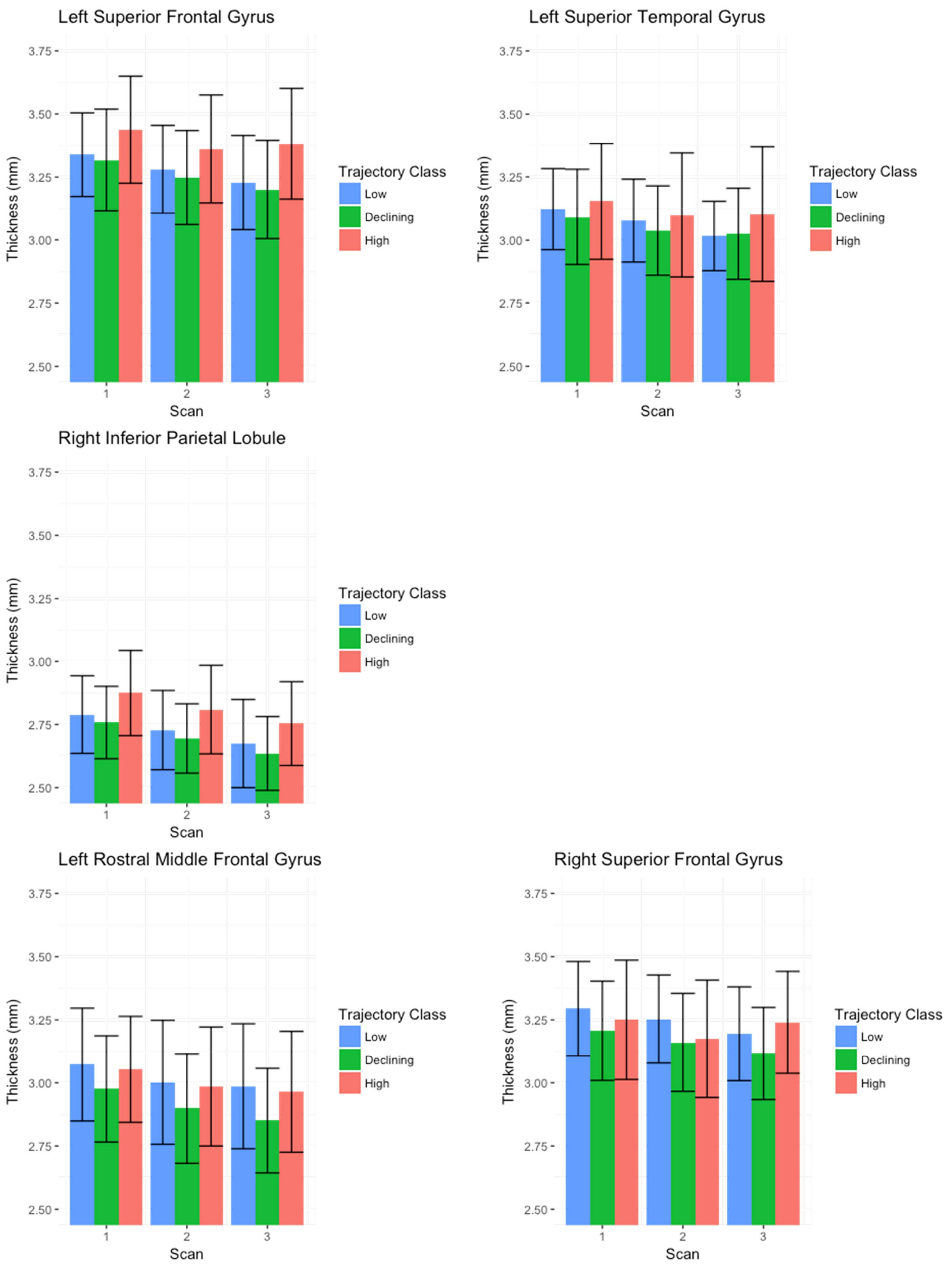

Figure S3. Mean cortical thickness from each cluster showing significant whole brain trajectory group differences in the temporal average of thickness is presented here, split by trajectory group and scan wave.

Note: Error bars indicate one standard deviation.