Longitudinal changes of ADHD symptoms in association with white matter microstructure: a tract-specific fixel-based analysis

Christienne G. Damatac[a,b], Sourena Soheili-Nezhad[a,b], Guilherme Blazquez Freches[a,b], Marcel P. Zwiers[a,b], Sanne de Bruijn[a,b], Seyma Ikde[a,b], Christel M. Portengen[a,b], Amy C. Abelmann[a,b], Janneke T. Dammers[a,b], Daan van Rooij[a,b], Sophie E. A. Akkermans[a,b], Jilly Naaijen[a,b], Barbara Franke[c,d], Jan K. Buitelaar[a,b,e], Christian F. Beckmann[a,b,f], Emma Sprooten[a,b]

[a] Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
[b] Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands
[c] Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands
[d] Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands
[e] Karakter Child and Adolescent Psychiatry University Centre, Nijmegen, The Netherlands
[f] Centre for Functional MRI of the Brain, Wellcome Centre for Integrative Neuroimaging, Nuffield Department of Clinical Neurosciences, University of Oxford, John Radcliffe Hospital, Oxford, UK

Correspondence to Christienne G. Damatac, Department of Cognitive Neuroscience, Donders Centre for Cognition, Donders Institute for Brain, Cognition and Behaviour, Kapittelweg 29, 6525 EN Nijmegen, The Netherlands; email at christienne.damatac@donders.ru.nl.

Abstract

Background: Variation in the longitudinal course of childhood attention deficit/hyperactivity disorder (ADHD) coincides with neurodevelopmental maturation of brain structure and function. Prior work has attempted to determine how alterations in white matter (WM) relate to changes in symptom severity, but much of that work has been done in smaller cross-sectional samples using voxel-based analyses. Using standard diffusion-weighted imaging (DWI) methods, we previously showed WM alterations were associated with ADHD symptom remission over time in a longitudinal sample of probands, siblings, and unaffected individuals. Here, we extend this work by further assessing the nature of these changes in WM microstructure by including an additional follow-up measurement (aged 18 – 34 years), and using the more physiologically informative fixel-based analysis (FBA). **Methods**: Data were obtained from 139 participants over 3 clinical and 2 follow-up DWI waves, and analyzed using FBA in regions-of-interest based on prior findings. We replicated previously reported significant models and extended them by adding another time-point, testing whether changes in combined ADHD and hyperactivity-impulsivity (HI) continuous symptom scores are associated with fixel metrics at follow-up. **Results**: Clinical improvement in HI symptoms over time was associated with more fiber density at follow-up in the left corticospinal tract (lCST) (t_{max}=1.092, standardized effect[SE]=0.044, p_{FWE}=0.016). Improvement in combined ADHD symptoms over time was associated with more fiber cross-section at follow-up in the lCST (t_{max}=3.775, SE=0.051, p_{FWE}=0.019). **Conclusions**: Aberrant white matter development involves both lCST micro- and macrostructural alterations, and its path may be moderated by preceding symptom trajectory.

Keywords: attention deficit hyperactivity disorder; magnetic resonance imaging; diffusion imaging; white matter; microstructure

Abbreviations: ADHD: attention-deficit hyperactivity disorder; **DTI**: diffusion tensor imaging; **DWI**: diffusion-weighted imaging; **FA**: fractional anisotropy; **FBA**: fixel-based analysis; **FC**: fiber cross-section; **FD**: fiber density; **FDC**: fiber density and cross-section; **FOD**: fiber orientation distribution; **HI**: hyperactivity-impulsivity; **IA**: inattention; **lCST**: left corticospinal tract; **lSLF**: left superior longitudinal fasciculus; **MRI**: magnetic resonance imaging; **SE**: standardized effect; **WM**: white matter

1. Introduction

Although (proto)typically considered a childhood syndrome, clinical trajectories of attention-deficit/hyperactivity disorder (ADHD) vary by individual. Many ADHD-affected adolescents exhibit improvement over time, but approximately two-thirds of them retain impairing symptoms into adulthood [1–3]. The neural substrates that determine this variable clinical course of childhood ADHD have been increasingly investigated through the years, yet the dynamic nature of these mechanisms in relation to maturation remains unclear. Theoretically, symptom remission occurs via brain compensation-reorganization, and/or normalization-convergence, with a possible fixed anomaly 'scar' or enduring neurological trait—all of which may concurrently arise in different brain regions [4]. In a double dissociative neurodevelopmental model of ADHD, the underlying neural mechanisms that control onset are distinct from those that drive remission [5]. Thus, onset can be characterized by dysfunctional subcortical structures remaining static throughout life, while remission may be separately associated with brain (particularly prefrontal cortex) maturation and compensation [4,6,7].

The theory that maturing frontal cortical regions compensate for initial childhood ADHD emergence via top-down regulatory processes, leading to eventual symptom remission, has been supported by magnetic resonance imaging (MRI) studies: Reduced symptom severity throughout development appears to correlate with prefrontal cortex maturation. White matter (WM) development in frontal-temporal areas subserving emotional and cognitive processes indeed continues to mature into early adulthood, coinciding with the typical age range of ADHD symptom remission [5,8–12]. Based on this model, it is possible to methodologically differentiate remitted from unaffected brains with MRI. Yet, previous neuroimaging studies have reported inconsistent results—perhaps because of study-specific differences (e.g. analysis methods, cross-sectional cohorts, sample characteristics). Considering this disorder's neurodevelopmental component, sample age is especially important, making systematic longitudinal studies essential in deconstructing the etiological timeline of brain mechanisms in reference to remission.

Diffusion-weighted imaging (DWI) is an *in vivo* MRI method which measures the magnitude and direction of water molecules diffusing through brain tissue, reflecting the underlying architecture of axons and their ensheathing myelin. Diffusion tensor imaging (DTI) has been the most commonly used DWI method in ADHD studies, which have usually reported tensor-derived, voxel-wise measures like fractional anisotropy (FA). One follow-up case-control DTI investigation in men suggested that ADHD is a lasting neurobiological trait irrespective of remission or persistence: Compared to those who did not have childhood ADHD, probands with both remittent and persistent ADHD showed widespread reduced FA three decades post-diagnosis [6]. Others showed that children who exhibited symptom improvement had the most FA anomalies at follow-up [13]. However, given ADHD's neurodevelopmental aspect, studying symptoms and brain tissues in late adolescence and early adulthood (as myelination continues) can give more relevant information about how remission is intertwined with maturation. While valuable, the few follow-up DTI reports to date were limited by categorical participant groups, sample characteristics (populations that were either pre-pubertal or well into adulthood), and only a single follow-up MRI measure—underscoring the need for more studies beyond the cross-sectional perspective.

A longitudinal design reveals temporal dynamics of underlying neurobiological processes and increases statistical power by reducing inter-subject variability. The NeuroIMAGE study and its latest follow-up, DELTA, is a longitudinal cohort of ADHD-affected probands, their siblings, and unaffected controls from childhood to adulthood [10,14–17]. We previously demonstrated that, at two different time-points and in two partly overlapping NeuroIMAGE samples, more improvement in combined ADHD and hyperactivity-impulsivity symptom scores over time were associated with lower FA at follow-up in an area where the left corticospinal tract (lCST) crosses the left superior longitudinal fasciculus (lSLF) [10,17]. In that same report, we also systematically demonstrated that symptom change is associated with neither baseline FA nor change in FA from baseline to follow-up. Symptom remission was counterintuitively and repeatedly associated with decreased FA later in life, whether from childhood to adolescence, or to early adulthood. Our longitudinal findings indicated divergent

80 WM microstructure trajectories between individuals with persistent and remittent symptoms at a
81 follow-up age range of 12 – 29 years. Now, we build on those previous results on the downstream
82 effect of symptom progression on WM microstructure by asking whether the same relationship exists
83 in the same brain areas at a later time window, when the cohort is aged 18 – 34 years (Figure 1).

84 Although *in vivo* WM microstructure has been most commonly studied through tensor-
85 derived metrics (e.g. FA), results from voxel-wise DTI-based methods can be unreliable or misleading
86 in areas with complex fiber architecture [18]. Given that tensor-based reconstructions are an average
87 across an entire voxel and that approximately 90% of voxels contain multiple fiber populations, we
88 applied a high angular diffusion model: constrained spherical deconvolution [19]. Voxel-based
89 methods that model crossing fibers (e.g. BEDPOSTX) only represent a subset of the full range of
90 possible fiber orientation distributions (FODs), whereas constrained spherical deconvolution
91 represents FODs as spherical harmonics, free to distinguish more or less arbitrary shapes [20]. Fixel-
92 based analysis (FBA) applies the constrained spherical deconvolution model and can more accurately
93 reconstruct a continuous FOD in both single- and multiple-fiber voxels—characterizing properties of
94 each "fixel," or specific fiber population in a voxel [21–25]. Fixels can be statistically analyzed for fiber-
95 specific indices of underlying physiology: fiber density (FD), a microstructural measure of the within-
96 voxel intra-axonal restricted compartment of a fiber population; fiber cross-section (FC), a
97 macrostructural measure of the area perpendicular to the fiber orientation; and fiber density and
98 cross-section (FDC), a combination of FD and FC [25]. Less FD can indicate axonal loss, while less FC
99 can indicate macroscopic fiber atrophy [26–28]. FBA resolves crossing fibers more accurately as well
100 as characterizes the microstructural and morphological, macrostructural properties of specific fiber
101 populations.

102 One cross-sectional FBA showed that ADHD-affected children who had reduced fine
103 motor competence also had lower WM microstructure in all three fixel metrics in the CST. These
104 results suggest that cases had fewer and/or thinner CST axons, which may lead to reduced fiber bundle
105 information transmission speed [29]. Despite the consistent clamor to resolve crossing fiber regions

and FBA's evident advantages, there have been no other published FBA applications in people with ADHD to our knowledge. Furthermore, besides our prior research, there have been no other longitudinal follow-up DWI studies of WM in the course of ADHD.

In an extension of our previous work in overlapping samples, here we followed 139 people over approximately 15 years. We used a more recent multi-shell DWI fiber model and a new follow-up measurement at an older age range. We aimed to assess the time-lag between the course of ADHD symptomatology and WM microstructure in *a priori* models and regions-of-interest. Because our smaller sample size at an older age range is not suitable for a data-driven search to discover any new relevant regions or tracts, the present analyses were intended to further understand the nature of our previous results. To compare FBA metrics to our previous FA findings, our first follow-up analysis used the same exact sample as our most recent longitudinal DTI study [17]. We hypothesized that, like in our earlier findings, ADHD symptom improvement would be associated with lower follow-up WM microstructure in the lSLF and lCST.

2. Methods

2.1 Participants

Clinical and MRI data were originally collected from probands with childhood ADHD, their first-degree relatives, and healthy families in one initial wave: NeuroIMAGE1 (W1) [14]. After an average of 3.7 years (standard deviation [SD] = 0.5 years), those participants were invited back for a second acquisition: NeuroIMAGE2 (W2). After a mean of 5.1 years (SD = 1.4 years), some individuals returned for another wave, DELTA (W3), which included only people who fulfilled full ADHD diagnostic criteria in at least one previous wave (Table 1). For the analyses here, we only included participants who had clinical data from at least two of the three waves and DWI data from W2 and/or W3 (Figure 1). For each time-point, there were no differences between the participants included in the current analyses and the complete sample in symptom severity, age, and sex ($p > 0.12$).

Given our longitudinal design, we did not split our participants into cases versus controls. Through the years, symptom scores and diagnoses varied through time and participant characteristics changed from wave to wave (Figure 2). Some individuals originally recruited as controls or unaffected siblings developed ADHD at a later time point and others recruited as ADHD participants remitted, further highlighting the complex, variable course of ADHD. Alternative to a case-control categorization, ADHD can be operationalized as a continuous trait [30,31]. In a previous cross-sectional study, we systematically showed that, compared to categorical diagnoses, continuous symptom measures are more sensitive to diffusion-weighted brain features in this sample [16]. Thus, all models here used ADHD symptom scores, optimally capturing the dynamic and continuous nature of the ADHD spectrum throughout development in this longitudinal cohort.

2.2 Clinical symptom measures

For continuous measures of symptom dimension severity and in accordance with our previous report, we used raw combined Conners' Parent Rating Scale (CPRS) scores from W1 and W2, and Conners' Adult ADHD Rating Scale (CAARS) scores from W3 for hyperactivity-impulsivity (HI) and inattention (IA) [17,32,33]. Here, we define symptom change (Δ) as the Conners' score difference: $\Delta score = score_{follow\text{-}up} - score_{baseline}$

Baseline versus follow-up scores were always positively correlated with each other (Figure S1). A more positive Δ value indicates the worsening of symptoms, while a more negative Δ value indicates the improvement of symptoms over time. In this report, we refer to "symptom remission" dimensionally and not diagnostically, i.e. a decrease or improvement in symptom severity over time.

At W1 and W2, we assessed history of comorbid disorders with the Kiddie Schedule for Affective Disorder and Schizophrenia Present and Lifetime Version (K-SADS-PL) semi-structured interview [34,35]. For children aged <12 years, the child's parents or the researchers assisted in completing the self-report questionnaires. Participants with elevated scores on ≥1 of the K-SADS-PL screening questions had to complete a full supplement for each disorder. At W3 (all participants were aged ≥18 years), we recorded history of comorbidity using the Structured Clinical Interview for DSM-5

Disorders (SCID-V) [36]. IQ was estimated using the vocabulary and block design subtests of the Wechsler Intelligence Scale for Children (WISC-III) or Wechsler Adult Intelligence Scale (WAIS-III). We excluded one whole dataset from a participant who had an estimated IQ <70. Our final sample's demographic characteristics are summarized in Table 1.

2.3 Diffusion-weighted imaging acquisition, pre-processing, and quality control

At W2, single-shell DWI data were acquired with a 1.5-Tesla AVANTO scanner (Siemens, Erlangen, Germany) equipped with an 8-channel receive-only phased-array head coil using the following parameters: echo time/repetition time (TE/TR) = 97/8500 ms; GRAPPA-acceleration factor 2; voxel size = 2 × 2 × 2.2 mm; *b*-values = 0 (5 volumes, interleaved) and 1000 (60 directions) s/mm^2; twice refocused pulsed-gradient spin-echo EPI; no partial Fourier. More details of this MRI data acquisition have been described previously [16,17]. Because our models only included follow-up neuroimaging data as a further investigation of the aforementioned analyses, we did not include W1 DWI.

At W3, multi-shell DWI data were acquired with a 3-Tesla Prisma scanner (Siemens, Erlangen, Germany) equipped with a 32-channel receive-only phased-array head coil using the following parameters: TE/TR = 75/2940 ms; multi-band acceleration factor = 3, voxel size = 1.8 mm^3; *b*-values = 0 (11 volumes, interleaved), 1250 (86 directions), and 2500 (85 directions) s/mm^2.

W2 and W3 images were pre-processed with MRtrix3 (version 3.0.1, http://www.mrtrix.org/) according to recommended FBA protocols for multi-shell data [24,25]. Pre-processing included denoising and unringing, motion and distortion correction, and bias field correction [37–43]. We visually inspected all corrected diffusion images and excluded whole datasets if any motion or distortion artefacts remained after pre-processing. After excluding 22 datasets, our final sample consisted of 154 total diffusion scans collected from 139 participants at W2 (N = 99) and W3 (N=55).

2.4 Fixel-based analysis

Following pre-processing, we computed two unique group average tissue response functions for W2: WM and cerebrospinal fluid (CSF) [44]. B0 images can be utilized like a second shell to estimate a CSF-specific response function for each participant [44]. By modeling distinct response functions for WM and CSF, we were able to enhance the signal from WM relative to CSF and include our single shell data in the multi-shell FBA pipeline. For W3, we calculate three response functions: WM, gray matter, and CSF [44]. We upsampled to 1.25 mm^3 and performed multi-shell multi-tissue constrained spherical deconvolution on all images, resulting in a WM fiber orientation distribution (FOD) within each voxel [22,45]. Afterwards, we performed joint bias field correction and global intensity normalization for each of the multi-tissue compartment parameters [41]. We then separately generated two study-specific FOD population templates for W2 and W3 using 40 unrelated participants from each wave per template. Symptom scores did not differ between the individuals included in the population templates, versus those of the overall samples ($P > 0.06$).

For each population template, we calculated the FD, log(FC), and FDC ($FDC = FD \cdot FC$) for each participant across all fixels. Instead of FC, we chose to calculate log(FC) so data would be centered around zero and normally distributed. The derivation of these fixel metrics, which are based on FOD lobe segmentation and subject-to-template registration warps, are described in detail elsewhere [46]. For each FOD template, we performed whole-brain fiber tractography and generated a fixel-fixel connectivity matrix from the whole-brain streamline tractograms.

To obtain each region-of-interest (left corticospinal tract: lCST; left superior longitudinal fasciculus: lSLFI, lSLFII, lSLFIII; right cingulum: rCG), we extracted the spherical harmonic peaks from each voxel of both FOD population templates. We then applied TractSeg, which is an automated convolutional neural network-based approach that directly segments tracts in fields of FOD peaks, circumventing any biases that may result from user-defined or atlas-based delineation [47]. Finally, we converted the resultant tractograms to fixel maps used as masks to constrain our search space during connectivity-based fixel enhancement [46] (Figure 3 and Figure S2).

2.5 Statistical analyses

To control for the lack of independence in our sample due to siblings, we designed multi-level exchangeability blocks per wave and used FSL PALM to generate a set of 5000 permutations per wave [48,49]. Our blocks did not allow permutation between all individuals; instead, we constrained permutations at both the whole-block level (i.e. permute between families of the same size) and within-block level (i.e. permute within families) (Figure S3). We used each set as an input for its respective wave to define permutations in data shuffling during nonparametric testing.

We demeaned our design matrices using *Jmisc* in *R* (version 4.0.2) and applied connectivity-based fixel enhancement to the fixel-fixel connectivity matrices using smoothed fixel data [46]. Using only models in which we previously found significant effects (i.e. not IA, but only HI and combined scores), for each fixel metric and each tract region-of-interest, we constructed general linear models (GLMs) to separately test whether combined ADHD or HI symptom score change (Δscore as independent variables) are associated with fixel metrics at follow-up (as dependent variables) [17]. Our covariates were: symptom score (either combined or HI) at baseline, change in age (Δage = age $_{\text{follow-up}}$ – age $_{\text{baseline}}$), age at baseline, sex, and head motion (framewise displacement) at follow-up. For the W2 FBA, follow-up was W2 and baseline was W1, while for the W3 FBA, follow-up was W3 and baseline was W2: fixel metric $_{\text{follow-up}}$ ~ Δscore + score $_{\text{baseline}}$ + Δage + age $_{\text{baseline}}$ + sex + head motion $_{\text{follow-up}}$

As a secondary cross-sectional analysis in W3 only, using the same aforementioned methods, we tested for an effect of CAARS score (IA, HI, and combined) on each of the three fixel metrics at the level of the whole brain, as well as rCG only. We used age, sex, and head motion (framewise displacement) as covariates: fixel metric ~ score + age + sex + head motion

Fixels were considered statistically significant at family-wise error-corrected *p* less than 0.05 ($p_{FWE} < 0.05$).

229 # 3. Results

230 ## 3.1 Association between white matter at Wave 2 and the change in symptoms from
231 Wave 1 to Wave 2

232 W2 fiber density (FD) in the lCST was significantly negatively associated with ΔHI score (t_{max} =
233 1.092, standardized effect [SE] = 0.044, p_{FWE} = 0.016; Figure 4). There were no other significant
234 associations between WM microstructure in any other tracts at follow-up and Δcombined or ΔHI score
235 (all $p_{FWE} \geq 0.051$; Table S1).

236 ## 3.2 Association between white matter at Wave 3 and the change in symptoms from
237 Wave 2 to Wave 3

238 W3 log of fiber cross-section (log[FC]) in the lCST was significantly negatively associated with
239 Δcombined symptom score (t_{max} = 3.775, SE = 0.051, p_{FWE} = 0.019; Figure 4). There were no other
240 significant associations between WM microstructure at follow-up in lSLFI, lSLFII, and lSLFIII and
241 Δcombined or ΔHI symptom score (all $p_{FWE} \geq 0.058$; Table S2).

242 ## 3.3 Association between white matter and symptoms at Wave 3 only

243 None of the global or rCG WM microstructure metrics were significantly associated with HI,
244 IA, or combined CAARS score cross-sectionally (all $p_{FWE} \geq 0.053$; Table S3).

245 # 4. Discussion

246 We conducted a unique study of WM microstructure and longitudinal ADHD symptom
247 development between ages 9 and 34 years. Using the FBA framework, we discovered two findings in
248 the lCST: (1) HI symptom improvement was associated with axonal expansion at follow-up, and (2)
249 combined ADHD symptom improvement was associated with a larger total cross-sectional area at
250 follow-up at a slightly later age-range. Initially, a previous voxel-wise analysis in an overlapping sample
251 found that improved HI symptoms were associated with lower follow-up FA (W1, aged 9 – 26 years)
252 [10]. Subsequently, we extended this sample by adding a second DWI time-point (W2, aged 12 – 29

253 years), and systematically applied and excluded specific models—ultimately replicating the same
254 effects on follow-up FA in the same WM region [17]. Given the counterintuitive nature of these
255 previous highly consistent results, the present analysis aimed to further understand the physiological
256 origins and its dynamic nature in relation to maturation. Thus here, in the exact same sample (W1-
257 W2) and including yet a third DWI acquisition (W3, aged 18 – 34 years), using the more advanced FBA
258 method, and employing the same GLMs in which we previously found significant voxel-wise effects,
259 we have found increased FD in relation to HI remission, and increased FC in relation to combined
260 symptom remission, in only the lCST and not the lSLF. In contrast to our previous finding using DTI-
261 based methods, our current finding using FBA is more intuitive in the direction of its effect: Indices
262 that are generally indicative of "stronger" fibers were associated with clinical improvements over
263 time.

264 The fixel metrics we used for quantifying WM microstructure contain complementary
265 information. FD is thought to be related to the microstructural properties of WM, whereas FC pertains
266 to the macrostructural properties (cross-sectional area). FD is an estimate of the intracellular volume
267 of fibers oriented in a particular direction. Higher FD at follow-up could result from developmental
268 processes like axon diameter growth, or more axons occupying a given space [50]. In our W1-to-W2
269 analysis, greater lCST axonal density in individuals who became less hyperactive-impulsive over time
270 suggests plasticity, or a greater ability to relay information, after symptom improvement.
271 Furthermore, higher myelin content could decrease the water exchange between intra- and extra-
272 axonal compartments, resulting in an apparent increase in the intra-axonal compartment's volume
273 and, hence, an increase in FD [42,51]. FC measures the morphological macroscopic change in the
274 cross-sectional area perpendicular to a fiber bundle (calculated during registration to the template
275 image). In W2-to-W3, higher follow-up lCST cross-sectional area in individuals whose combined
276 symptom score improved, again, suggests plasticity, greater myelination, or fiber bundle organization
277 after symptom remission [25].

278 Although the direction of effects in our FBA analyses are opposite to that of our
279 aforementioned FA analyses, they are not incompatible. In some cases, crossing fiber complexity can
280 have an inverse correlation with FA, wherein greater complexity occurs when more fixels in a voxel
281 have the same fiber density [52]. An analogous inverse association exists in our previous W1-to-W2
282 voxel-wise analysis, wherein less follow-up FA was associated with improved HI symptom score.
283 Notably, our results were in the approximate location of where the lSLF and lCST cross, while our
284 present fixel-wise results in the lCST seem to be absent from where the lSLF crosses this tract.
285 Therefore, as we previously suggested, our tract-based spatial statistics results may have been due to
286 the neuroanatomical location of the effects, which, when labeled with an atlas, were in an area where
287 these tracts cross. Compared to our voxel-wise study, we presently accounted for crossing fibers
288 better through FBA, as well as the specific, FOD-based segmentation of these tracts as separate
289 regions-of-interest. Accordingly, symptom improvement over time can conceivably be associated with
290 increased CST fiber maturation, which by our previous DTI methods may have appeared as reduced
291 FA in voxels where a more dominant SLF crosses those corticospinal fibers.

292 The lCST is the only tract in which we have consistently found longitudinal effects. However,
293 a cross-sectional study of symptoms in an overlapping W1 sample found the most FA differences in
294 the right cingulum-angular bundle [16]. Anatomically, this differs from that of the present and
295 previous longitudinal effects, which suggests a dissociation in the WM tracts associated with cross-
296 sectional differences versus those that are associated with symptom remission. This dissociation
297 points to the neurodevelopmental models of remission, which all predict atypical neural features in
298 adults with persistent ADHD, but have different predictions about those with remittent symptoms: If
299 symptom remission occurred via WM normalization, convergence, or passive delayed maturation,
300 then we would have observed neurological alterations at follow-up in participants with persistent
301 symptoms, but no differences between those with remittent symptoms and healthy controls. If,
302 regardless of symptom trajectory, this disorder imparted an indelible mark or scar on the brain, then
303 we would not have observed follow-up neurological differences between those with persistent versus

304 remittent symptoms, and only healthy controls would have been differentiated by our follow-up
305 analyses. If symptom remission occurred via compensation or reorganization, then remitted brains
306 would have differed from both the never affected and the persistent ADHD brains, but in different
307 ways. The dissociation we observed in the tract that is important for symptom remission versus the
308 tract that is important for symptom severity implies the last model of remission. According to our
309 current findings, we tentatively suggest an interpretation consistent with our previous report: In a
310 top-down fashion, remitters may have learned compensatory strategies to overcome symptoms as
311 they aged, while persisters may have either learned disadvantageous strategies, other beneficial (but
312 insufficiently effective) compensatory strategies, or none at all—leading to diverging WM
313 development trajectories in specific brain regions in individuals with persistent ADHD symptoms
314 (Figure S4).

315 Based on our longitudinal design, we postulate that different WM alteration patterns are
316 associated with symptom trajectory in a tract-specific manner: HI symptom remission preceded lCST
317 plasticity at 20 years median age (range 12 – 29 years), and combined symptom remission preceded
318 lCST plasticity at 26 years median age (range 18 – 34 years). Perhaps our sample at a slightly younger
319 age, in response to HI symptom improvement or learning new skills, gained more lCST fibers over
320 time. Tract expansion could have been a compensatory mechanism to improve motor control,
321 followed by more myelination of those fibers. Then, as our participants became slightly older,
322 improvement in both dimensions may have led to greater lCST WM microstructure and improved
323 motor control (Figure S5). We can speculate that improved IA (and related executive control) could
324 help suppress HI, leading to greater motor control evinced as larger FC at a later age. In our remitters,
325 higher measures of lCST WM might also result from reorganization in other brain areas outside of the
326 tracts we studied. Along with our previous study, we have again found that alterations in WM
327 microstructure appear to follow symptom improvement. Speculatively, this suggests that WM
328 changes may be a downstream result of ADHD symptom remission.

329 A strength of the current study is its large sample size over three clinical and two DWI time-
330 points. Our approach using two separate follow-up analyses lent further characterization to the
331 temporal dynamics of ADHD-WM microstructure interplay. Of particular concern given this disorder,
332 we mitigated potential confounding effects of head motion through careful data screening, correction
333 during preprocessing, and inclusion as a covariate in our models. Using multi-shell FBA, we
334 demonstrated that WM-associated differences are fiber-specific even within regions of crossing fibers,
335 and we were able to further characterize WM micro- and macrostructural properties. Nonetheless, a
336 limitation of this study is that the first follow-up analysis included DWI data acquired with only one
337 relatively low b-value at 1.5T, which may have precluded us from discovering effects in other tracts
338 and/or symptom dimensions since higher b-value shells improve correspondence between FD
339 estimates and intra-axonal signal fraction simulations by increasing extra-axonal signal suppression
340 [53]. Second, our inability to replicate our previous cross-sectional tractography findings may stem
341 from the different DWI analysis methods used; also, W3 included a comparatively smaller number of
342 affected-only individuals, which reduced our statistical power to detect smaller cross-sectional effects
343 in the cingulum. Third, our follow-up samples were prone to selection bias from attrition and our
344 explicit selection criteria in W3, where we also used different instruments (CRPS vs. CAARS) and raters.
345 Returning participants were different from those who participated only once. Finally, even in a
346 longitudinal study, we cannot prove causality. ADHD symptom persistence is likely associated with
347 many other factors in daily life, or medication, or comorbid symptomatology—and any combination
348 of these could also contribute to neurological differences at follow-up. Given our small sample size,
349 we used *a priori* regions-of-interest and models, but a larger whole-brain analysis would be a more
350 sufficient test of which mechanisms of remission are at play.

351 Our findings contribute to the growing body of evidence describing the progression of
352 symptoms in relation to WM development. Defining the correlates and predictors of remission may
353 eventually lead to an improved allocation of treatment resources for persistent or complicated ADHD.

A better understanding of the underlying neural mechanisms of these changes in time can contribute to the promotion of favorable future perspectives for children and adolescents with this disorder.

5. Funding, acknowledgments, and financial disclosures

The authors would like to thank all of the families who participated in this study and all of the researchers who collected the data. This study sample is from the NeuroIMAGE project. NeuroIMAGE is the longitudinal follow-up study of the Dutch part of the International Multisite ADHD Genetics (IMAGE) project, which was a multi-site, international effort. NeuroIMAGE was supported by an NWO Large Investment Grant 1750102007010 and NWO Brain & Cognition an Integrative Approach Grant (433-09-242) (to J.K.B.), and grants from Radboud University Nijmegen Medical Center, University Medical Center Groningen and Accare, and VU University Amsterdam. Funding agencies had no role in study design, data collection, interpretation or influence on writing. J.N. is supported by an NWO Veni grant (no. VI.Veni.194.032). B.F. has received educational speaking fees from Medice. J.K.B. has been in the past 3 years a consultant to / member of advisory board of / and/or speaker for Takeda/Shire, Roche, Medice, Angelini, Janssen, and Servier. He is not an employee of any of these companies, and not a stock shareholder of any of these companies. He has no other financial or material support, including expert testimony, patents, royalties. All other authors report no biomedical financial interests or potential conflicts of interest.

6. References

[1] Faraone, S. V., Biederman, J. and Mick, E. (2006) The age-dependent decline of attention deficit hyperactivity disorder: A meta-analysis of follow-up studies. *Psychological Medicine*, **36**, 159–65. https://doi.org/10.1017/S003329170500471X

[2] Faraone, S. V, Asherson, P., Banaschewski, T., Biederman, J., Buitelaar, J.K., Ramos-Quiroga, J.A. et al. (2015) Attention-deficit/hyperactivity disorder. *Nature Reviews Disease Primers*, Radboud University Medical Center, Donders Institute for Brain, Cognition and Behaviour,

379 Departments of Human Genetics and Psychiatry, Nijmegen, The Netherlands. **1**, 15020.
380 https://doi.org/10.1038/nrdp.2015.20
381 [3] Sibley, M.H., Mitchell, J.T. and Becker, S.P. (2016) Method of adult diagnosis influences
382 estimated persistence of childhood ADHD: a systematic review of longitudinal studies. *The*
383 *Lancet Psychiatry*, Elsevier Ltd. **3**, 1157–65. https://doi.org/10.1016/S2215-0366(16)30190-0
384 [4] Sudre, G., Mangalmurti, A. and Shaw, P. (2018) Growing out of attention deficit hyperactivity
385 disorder: Insights from the 'remitted' brain. *Neuroscience and Biobehavioral Reviews*, **94**, 198–
386 209. https://doi.org/10.1016/j.neubiorev.2018.08.010
387 [5] Halperin, J.M. and Schulz, K.P. (2006) Revisiting the role of the prefrontal cortex in the
388 pathophysiology of attention-deficit/hyperactivity disorder. *Psychological Bulletin*,
389 Department of Psychology, Queens College of the City University of New York, Flushing, NY
390 11367, USA. jeffrey.halperin@qc.cuny.edu. **132**, 560–81. https://doi.org/10.1037/0033-
391 2909.132.4.560
392 [6] Cortese, S, Klein, R.G., Mannuzza, S., Ramos-olazagasti, Maria A, Castellanos, F., Cortese, S.,
393 Imperati, D., Zhou, J. et al. (2013) White Matter Alterations at 33-Year Follow-Up in Adults with
394 Childhood Attention-Deficit/Hyperactivity Disorder. *Biol Psychiatry*, **74**, 591–8.
395 https://doi.org/10.1016/j.biopsych.2013.02.025.White
396 [7] Shaw, P., Sudre, G., Wharton, A., Weingart, D., Sharp, W. and Sarlls, J. (2015) White matter
397 microstructure and the variable adult outcome of childhood attention deficit hyperactivity
398 disorder. *Neuropsychopharmacology*, Nature Publishing Group. **40**, 746–54.
399 https://doi.org/10.1038/npp.2014.241
400 [8] Clerkin, S.M., Schulz, K.P., Berwid, O.G., Fan, J., Newcorn, J.H., Tang, C.Y. et al. (2013) Thalamo-
401 cortical activation and connectivity during response preparation in adults with persistent and
402 remitted ADHD. *American Journal of Psychiatry*, **170**, 1011–9.
403 https://doi.org/10.1176/appi.ajp.2013.12070880
404 [9] Shaw, P., Malek, M., Watson, B., Greenstein, D., De Rossi, P. and Sharp, W. (2013) Trajectories

of cerebral cortical development in childhood and adolescence and adult attention-deficit/hyperactivity disorder. *Biological Psychiatry*, Elsevier. **74**, 599–606. https://doi.org/10.1016/j.biopsych.2013.04.007

[10] Francx, W., Zwiers, M.P., Mennes, M., Oosterlaan, J., Heslenfeld, D., Hoekstra, P.J. et al. (2015) White matter microstructure and developmental improvement of hyperactive/impulsive symptoms in attention-deficit/hyperactivity disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, Karakter Child and Adolescent Psychiatry University Centre, Nijmegen, The Netherlands. **56**, 1289–97. https://doi.org/10.1111/jcpp.12379

[11] Cheung, C.H.M., Rijdijk, F., McLoughlin, G., Faraone, S. V, Asherson, P. and Kuntsi, J. (2015) Childhood predictors of adolescent and young adult outcome in ADHD. *Journal of Psychiatric Research*, **62**, 92–100. https://doi.org/10.1016/j.jpsychires.2015.01.011

[12] Lebel, C. and Deoni, S. (2018) The development of brain white matter microstructure. *NeuroImage*, **182**, 207–18. https://doi.org/10.1016/j.neuroimage.2017.12.097

[13] Sudre, G., Sharp, W., Kundzicz, P., Bouyssi-Kobar, M., Norman, L., Choudhury, S. et al. (2020) Predicting the course of ADHD symptoms through the integration of childhood genomic, neural, and cognitive features. *Molecular Psychiatry*, Springer US. https://doi.org/10.1038/s41380-020-00941-x

[14] von Rhein, D., Mennes, M., van Ewijk, H., Groenman, A.P., Zwiers, M.P., Oosterlaan, J. et al. (2015) The NeuroIMAGE study: a prospective phenotypic, cognitive, genetic and MRI study in children with attention-deficit/hyperactivity disorder. Design and descriptives. *European Child and Adolescent Psychiatry*, Centre for Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, The Netherlands, Daniel.vonRhein@gmail.com. **24**, 265–81. https://doi.org/10.1007/s00787-014-0573-4

[15] van Ewijk, H., Heslenfeld, D.J., Zwiers, M.P., Faraone, S. V., Luman, M., Hartman, C.A. et al. (2014) Different mechanisms of white matter abnormalities in attention-deficit/ hyperactivity disorder: A diffusion tensor imaging study. *Journal of the American Academy of Child and*

Adolescent Psychiatry, Elsevier Inc. **53**, 790-799.e3. https://doi.org/10.1016/j.jaac.2014.05.001

[16] Damatac, C.G., Chauvin, R.J.M., Zwiers, M.P., van Rooij, D., Akkermans, S.E.A., Naaijen, J. et al. (2020) White Matter Microstructure in Attention-Deficit/Hyperactivity Disorder: A Systematic Tractography Study in 654 Individuals. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, Elsevier Inc. 1–10. https://doi.org/10.1016/j.bpsc.2020.07.015

[17] Leenders, A. E. M. & Damatac, C.G., Soheili-Nezhad, S., Chauvin, R.J.M., Mennes, M.J.J., Zwiers, M.P., Rooij, D. et al. (2021) Associations between attention-deficit hyperactivity disorder (ADHD) symptom remission and white matter microstructure: A longitudinal analysis. *JCPP Advances*, 1–11. https://doi.org/10.1002/jcv2.12040

[18] Seehaus, A., Roebroeck, A., Bastiani, M., Fonseca, L., Bratzke, H., Lori, N. et al. (2015) Histological validation of high-resolution DTI in human post mortem tissue. *Frontiers in Neuroanatomy*, **9**. https://doi.org/10.3389/fnana.2015.00098

[19] Jeurissen, B., Leemans, A., Tournier, J.D., Jones, D.K. and Sijbers, J. (2013) Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Human Brain Mapping*, **34**, 2747–66. https://doi.org/10.1002/hbm.22099

[20] Tournier, J.D., Calamante, F., Gadian, D.G. and Connelly, A. (2004) Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, **23**, 1176–85. https://doi.org/10.1016/j.neuroimage.2004.07.037

[21] Schilling, K.G., Janve, V., Gao, Y., Stepniewska, I., Landman, B.A. and Anderson, A.W. (2018) Histological validation of diffusion MRI fiber orientation distributions and dispersion. *NeuroImage*, **165**, 200–21. https://doi.org/10.1016/j.neuroimage.2017.10.046

[22] Tournier, J.-D., Calamante, F. and Connelly, A. (2007) Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage*, **35**, 1459–72. https://doi.org/https://doi.org/10.1016/j.neuroimage.2007.02.016

[23] Tournier, J.-D., Calamante, F. and Connelly, A. (2013) Determination of the appropriate b value and number of gradient directions for high-angular-resolution diffusion-weighted imaging. *NMR in Biomedicine*, England. **26**, 1775–86. https://doi.org/10.1002/nbm.3017

[24] Tournier, J.D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M. et al. (2019) MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage*, Elsevier Ltd. **202**, 116137. https://doi.org/10.1016/j.neuroimage.2019.116137

[25] Raffelt, D.A., Tournier, J.-D.D., Smith, R.E., Vaughan, D.N., Jackson, G., Ridgway, G.R. et al. (2017) Investigating white matter fibre density and morphology using fixel-based analysis. *NeuroImage*, Elsevier. **144**, 58–73. https://doi.org/10.1016/j.neuroimage.2016.09.029

[26] Gajamange, S., Raffelt, D., Dhollander, T., Lui, E., van der Walt, A., Kilpatrick, T. et al. (2018) Fibre-specific white matter changes in multiple sclerosis patients with optic neuritis. *NeuroImage: Clinical*, **17**, 60–8. https://doi.org/10.1016/j.nicl.2017.09.027

[27] Mito, R., Raffelt, D., Dhollander, T., Vaughan, D.N., Tournier, J.D., Salvado, O. et al. (2018) Fibre-specific white matter reductions in Alzheimer's disease and mild cognitive impairment. *Brain*, **141**, 888–902. https://doi.org/10.1093/brain/awx355

[28] Rojas-Vite, G., Coronado-Leija, R., Narvaez-Delgado, O., Ramírez-Manzanares, A., Marroquín, J.L., Noguez-Imm, R. et al. (2019) Histological validation of per-bundle water diffusion metrics within a region of fiber crossing following axonal degeneration. *NeuroImage*, **201**, 116013. https://doi.org/https://doi.org/10.1016/j.neuroimage.2019.116013

[29] Hyde, C., Fuelscher, I., Sciberras, E., Efron, D., Anderson, V.A. and Silk, T. (2020) Understanding motor difficulties in children with ADHD: A fixel-based analysis of the corticospinal tract. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, Elsevier. **105**, 110125. https://doi.org/10.1016/j.pnpbp.2020.110125

[30] Lahey, B.B. and Willcutt, E.G. (2010) Predictive Validity of a Continuous Alternative to Nominal Subtypes of Attention-Deficit/Hyperactivity Disorder for DSM – V. *Journal of Clinical Child &*

Adolescent Psychology, **39**, 761–75. https://doi.org/10.1080/15374416.2010.517173

[31] Marcus, D.K. and Barry, T.D. (2011) Does attention-deficit/hyperactivity disorder have a dimensional latent structure? A taxometric analysis. *Journal of Abnormal Psychology*, **120**, 427–42. https://doi.org/10.1037/a0021405

[32] Conners, C.K., Sitarenios, G., Parker, J.D.A. and Epstein, J.N. (1998) The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology*, **26**, 257–68. https://doi.org/10.1023/a:1022602400621

[33] Conners, C.K., Erhardt, D., Epstein, J.N., Parker, J.D.A., Sitarenios, G. and Sparrow, E. (1999) Self-ratings of ADHD symptoms in adults I: Factor structure and normative data. *Journal of Attention Disorders*,. https://doi.org/10.1177/108705479900300303

[34] Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P. et al. (1997) Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry*, United States. **36**, 980–8. https://doi.org/10.1097/00004583-199707000-00021

[35] Donker, T., Comijs, H., Cuijpers, P., Terluin, B., Nolen, W., Zitman, F. et al. (2010) The validity of the Dutch K10 and extended K10 screening scales for depressive and anxiety disorders. *Psychiatry Research*,. https://doi.org/10.1016/j.psychres.2009.01.012

[36] First, M.B., Williams, J.B.W., Karg, R.S., Spitzer, R.L., Arntz, A.R. and Association, A.P. (2018) SCID-5-S: Gestructureerd Klinisch Interview Voor DSM-5 Syndroomstoornissen [Internet]. Boom.

[37] Zwiers, M.P. (2010) NeuroImage Patching cardiac and head motion artefacts in diffusion-weighted images. *NeuroImage*, Elsevier Inc. **53**, 565–75. https://doi.org/10.1016/j.neuroimage.2010.06.014

[38] Veraart, J., Novikov, D.S., Christiaens, D., Ades-aron, B., Sijbers, J. and Fieremans, E. (2016) Denoising of diffusion MRI using random matrix theory. *NeuroImage*,.

https://doi.org/10.1016/j.neuroimage.2016.08.016

[39] Andersson, J.L.R. and Sotiropoulos, S.N. (2016) An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*,. https://doi.org/10.1016/j.neuroimage.2015.10.019

[40] Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A. et al. (2010) N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*,. https://doi.org/10.1109/TMI.2010.2046908

[41] Raffelt, D., Dhollander, T., Tournier, J.D., Tabbara, R., Smith, R.E., Pierre, E. et al. (2017) Bias field correction and intensity normalisation for quantitative analysis of apparent fiber density. *Proc Intl Soc Mag Reson Med*,.

[42] Raffelt, D., Tournier, J.D., Rose, S., Ridgway, G.R., Henderson, R., Crozier, S. et al. (2012) Apparent Fibre Density: A novel measure for the analysis of diffusion-weighted magnetic resonance images. *NeuroImage*, Elsevier Inc. **59**, 3976–94. https://doi.org/10.1016/j.neuroimage.2011.10.045

[43] Kellner, E., Dhital, B., Kiselev, V.G. and Reisert, M. (2016) Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magnetic Resonance in Medicine*,. https://doi.org/10.1002/mrm.26054

[44] Dhollander, T., Raffelt, D. and Connelly, A. (2016) Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion MR data without a co-registered T1 image.

[45] Jeurissen, B., Tournier, J.D., Dhollander, T., Connelly, A. and Sijbers, J. (2014) Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage*, Elsevier Inc. **103**, 411–26. https://doi.org/10.1016/j.neuroimage.2014.07.061

[46] Raffelt, D.A., Smith, R.E., Ridgway, G.R., Tournier, J.D., Vaughan, D.N., Rose, S. et al. (2015) Connectivity-based fixel enhancement: Whole-brain statistical analysis of diffusion MRI measures in the presence of crossing fibres. *NeuroImage*, Elsevier B.V. **117**, 40–55. https://doi.org/10.1016/j.neuroimage.2015.05.039

[47] Wasserthal, J., Neher, P. and Maier-Hein, K.H. (2018) TractSeg - Fast and accurate white matter tract segmentation. *NeuroImage*, Elsevier Ltd. **183**, 239–53. https://doi.org/10.1016/j.neuroimage.2018.07.070

[48] Winkler, A.M., Webster, M.A., Vidaurre, D., Nichols, T.E. and Smith, S.M. (2015) Multi-level block permutation. *NeuroImage*, The Authors. **123**, 253–68. https://doi.org/10.1016/j.neuroimage.2015.05.092

[49] Winkler, A.M., Ridgway, G.R., Webster, M.A., Smith, S.M. and Nichols, T.E. (2014) Permutation inference for the general linear model. *NeuroImage*, The Authors. **92**, 381–97. https://doi.org/10.1016/j.neuroimage.2014.01.060

[50] Genc, S., Smith, R.E., Malpas, C.B., Anderson, V., Nicholson, J.M., Efron, D. et al. (2018) Development of white matter fibre density and morphology over childhood: A longitudinal fixel-based analysis. *NeuroImage*, Elsevier Ltd. **183**, 666–76. https://doi.org/10.1016/j.neuroimage.2018.08.043

[51] Pannek, K., Fripp, J., George, J.M., Fiori, S., Colditz, P.B., Boyd, R.N. et al. (2018) Fixel-based analysis reveals alterations is brain microstructure and macrostructure of preterm-born infants at term equivalent age. *NeuroImage: Clinical*, Elsevier. **18**, 51–9. https://doi.org/10.1016/j.nicl.2018.01.003

[52] Grazioplene, R.G., Bearden, C.E., Subotnik, K.L., Ventura, J., Haut, K., Nuechterlein, K.H. et al. (2018) Connectivity-enhanced diffusion analysis reveals white matter density disruptions in first episode and chronic schizophrenia. *NeuroImage: Clinical*, Elsevier. **18**, 608–16. https://doi.org/10.1016/j.nicl.2018.02.015

[53] Genc, S., Tax, C.M.W., Raven, E.P., Chamberland, M., Parker, G.D. and Jones, D.K. (2020) Impact of b-value on estimates of apparent fibre density. *Human Brain Mapping*, **41**, 2583–95. https://doi.org/10.1002/hbm.24964

7. Tables and Figures

	Wave 1		Wave 2		Wave 3			
	N = 119		N = 135		N = 55			
	Mean	**(SD)**	**Mean**	**(SD)**	**Mean**	**(SD)**	**Test Statistic**	**(*p*)**
Age, years	16.98	(3.47)	20.22	(3.48)	24.82	(4.07)		
Sex, female	N = 48	40%	N = 53	39%	N = 19	35%	$\chi^2_3 = 1.51$	(0.68)
Estimated IQ[a]	98.41	(14.81)	105	(16.82)	102.35	(14.17)	$F_{2,192} = 2.23$	(0.11)
Head motion, framewise displacement	0.47	(0.17)	0.53	(0.42)	0.82	(0.17)	$F_{2,298} = 24.71$	($<10^{-10}$)
Handedness, right	N = 96	81%	N = 111	82%	N = 45	82%	$\chi^2_3 = 4.61$	(0.20)
Medication ever used, yes[b,c]	N = 66	56%	N = 45	33%	N = 42	76%	$\chi^2_3 = 32.03$	($<10^{-6}$)
Symptom raw score by diagnostic group[d]								
Combined score[e]	13.03	(12.33)	12.96	(13.05)	20.53	(9.34)		
Hyperactivity-impulsivity score	4.94	(5.73)	4.71	(5.68)	10.85	(5.38)		
Inattention score	8.05	(7.51)	7.55	(7.45)	9.67	(4.85)		
Comorbidity diagnosis								
Anxiety disorder	N = 2	2%	N = 2	1%	N = 1	2%	$\chi^2_3 = 57.344$	($<10^{-11}$)
Avoidant personality disorder	N = 1	1%	N = 2	1%	N = 0	0%	$\chi^2_3 = 81.886$	($<10^{-15}$)
Conduct disorder	N = 2	2%	N = 2	1%	*NA*	*NA*	$\chi^2_2 = 55.107$	($<10^{-11}$)
Major depression	N = 1	1%	N = 1	1%	N = 4	7%	$\chi^2_3 = 57.435$	($<10^{-11}$)
Oppositional defiant disorder	N = 15	13%	N = 16	12%	*NA*	*NA*	$\chi^2_2 = 7.8901$	(7.89)
Panic disorder	N = 0	0%	N = 1	1%	N = 0	0%	$\chi^2_3 = 137.76$	($<10^{-15}$)
Substance use								
Alcohol	N = 19	16%	N = 21	16%	N = 7	13%	$\chi^2_3 = 6.2621$	(0.10)
Tobacco	N = 43	36%	N = 45	33%	N = 12	22%	$\chi^2_3 = 5.1952$	(0.16)
Cannabis or hash	N = 22	18%	N = 23	17%	N = 6	11%	$\chi^2_3 = 6.8817$	(0.08)
Other drugs	N = 6	5%	N = 6	4%	N = 2	4%	$\chi^2_3 = 23.734$	($<10^{-4}$)

565 **Table 1.** Demographic and clinical characteristics of participants at Wave 1 (W1), Wave 2 (W2), and
566 Wave 3 (W3) with mean and standard deviation (or numerical count and percentage). W3 included
567 only those who fulfilled full ADHD diagnostic criteria in at least one previous wave. Values reported
568 here are for all participants in the final sample after all quality control (N = 139).
569 [a] IQ was estimated using the vocabulary and block design subtests of the Wechsler Intelligence Scale
570 for Children or Wechsler Adult Intelligence Scale.
571 [b] Medication ever used: Whether or not participants had ever taken ADHD medication.
572 [c] Medications: Ritalin (methylphenidate), Concerta (methylphenidate), Strattera (atomoxetine), and
573 any other ADHD medication. The majority of patients were taking prescription medication for ADHD,
574 mostly methylphenidate or atomoxetine.
575 [d] Symptom scores in W1 and W2 were collected via Conners' Parent Rating Scale, and W3 scores were
576 collected via the Conners' Adult ADHD Rating Scale.
577 [e] Combined symptom score: Sum of hyperactivity-impulsivity and inattention scores.
578

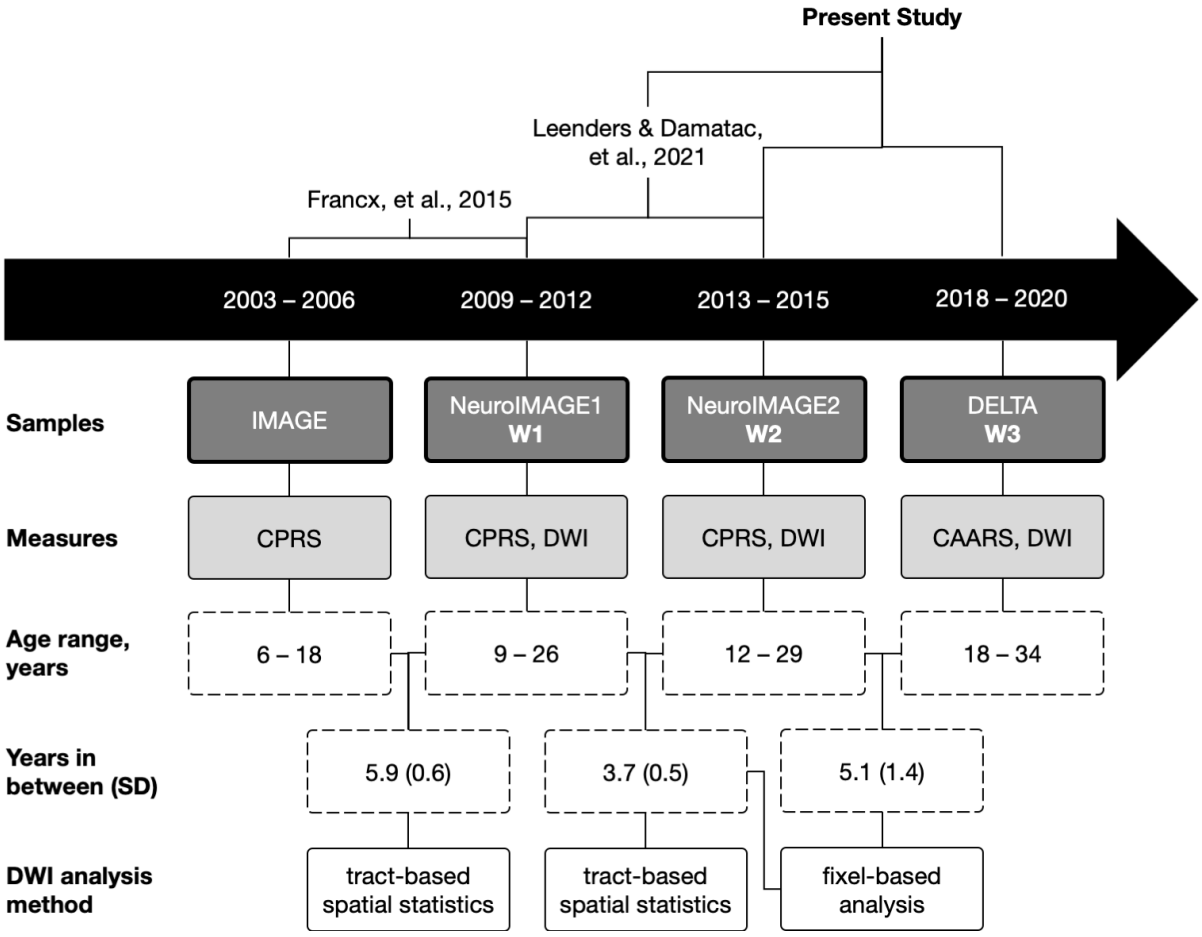

Figure 1. Schematic of how this study chronologically relates to previous studies, the samples included in each, relevant clinical and neuroimaging measurements, study sample age ranges, mean years (standard deviation) in between each acquisition wave, and the analysis methods used. The present study is a fixel-based analysis of W1 to W2 and W2 to W3, using only the models in which we found significant effects in a previous voxel-wise tract-based spatial statistical analysis of W1 to W2.

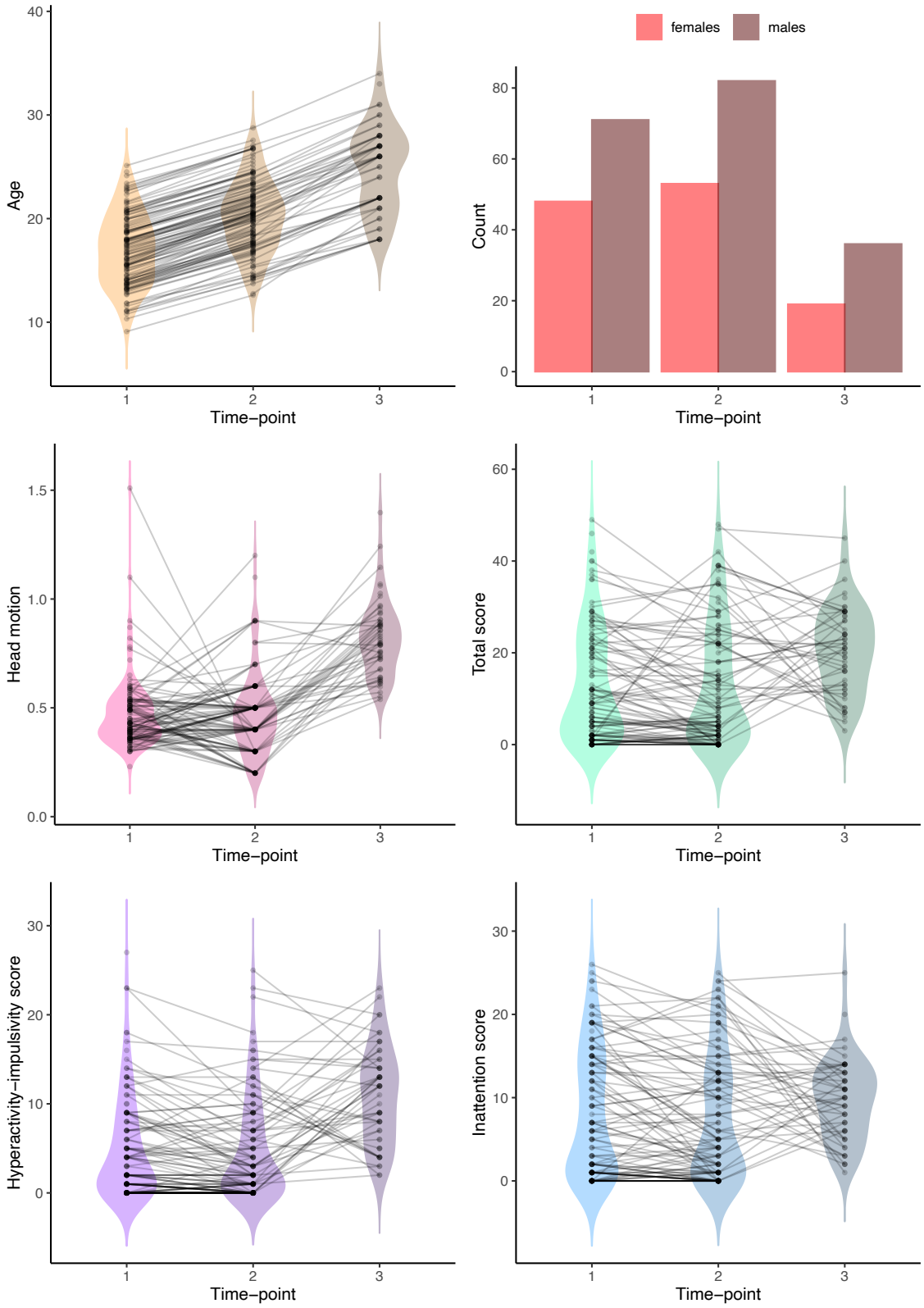

Figure 2. Change in participant characteristics from Wave 1, to Wave 2, to Wave 3. Longitudinal data points are connected by a line. Note that participants at W3 were selected on the basis of their history of ADHD diagnosis, so W3 tends to differ quite markedly from the other two waves, which also include never-affected controls. This conceals the typical pattern of average symptom remission that would be expected in a follow-up study without this selection criterion.

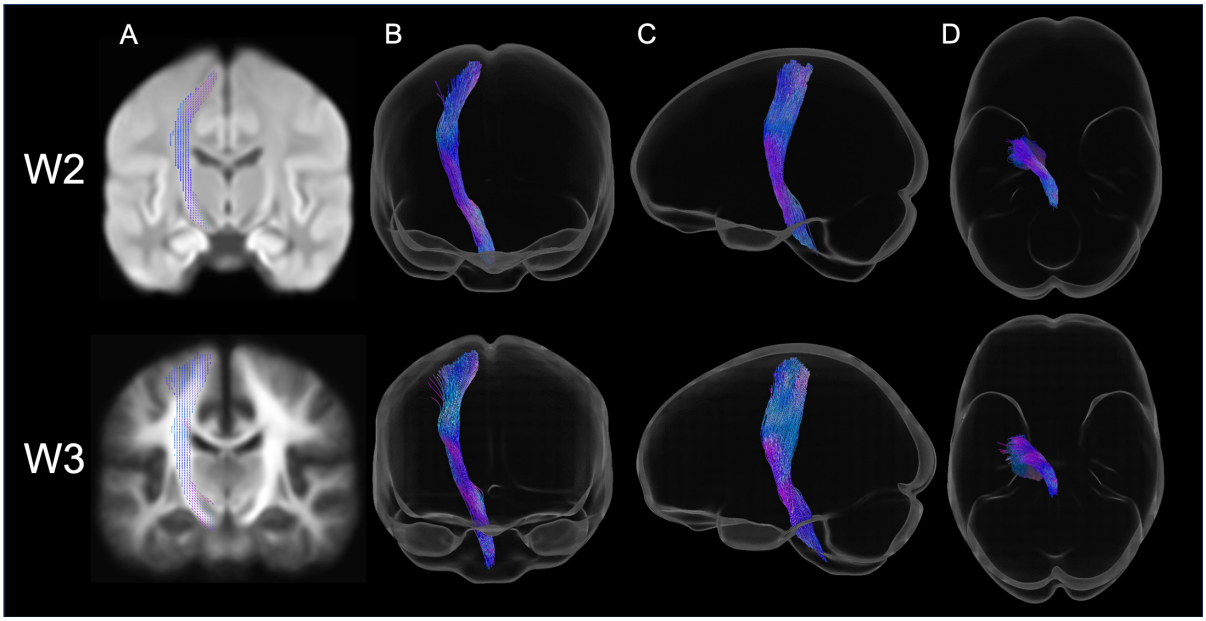

Figure 3. Tract-specific region-of-interest masks of the left corticospinal tract for Wave 2 (top) and Wave 3 (bottom) colored by direction (red: left-right, green: anterior-posterior, blue: inferior-superior). (A) Fixel mask overlaid on a single representative coronal slice of the study-specific white matter fiber orientation distribution template. Template contrast was adjusted and fixels have been thickened for visualization. (B) Coronal, (C) sagittal, and (D) axial views of the tract reconstruction from TractSeg (applied to each fiber orientation distribution template) and displayed in glass brains for visualization.

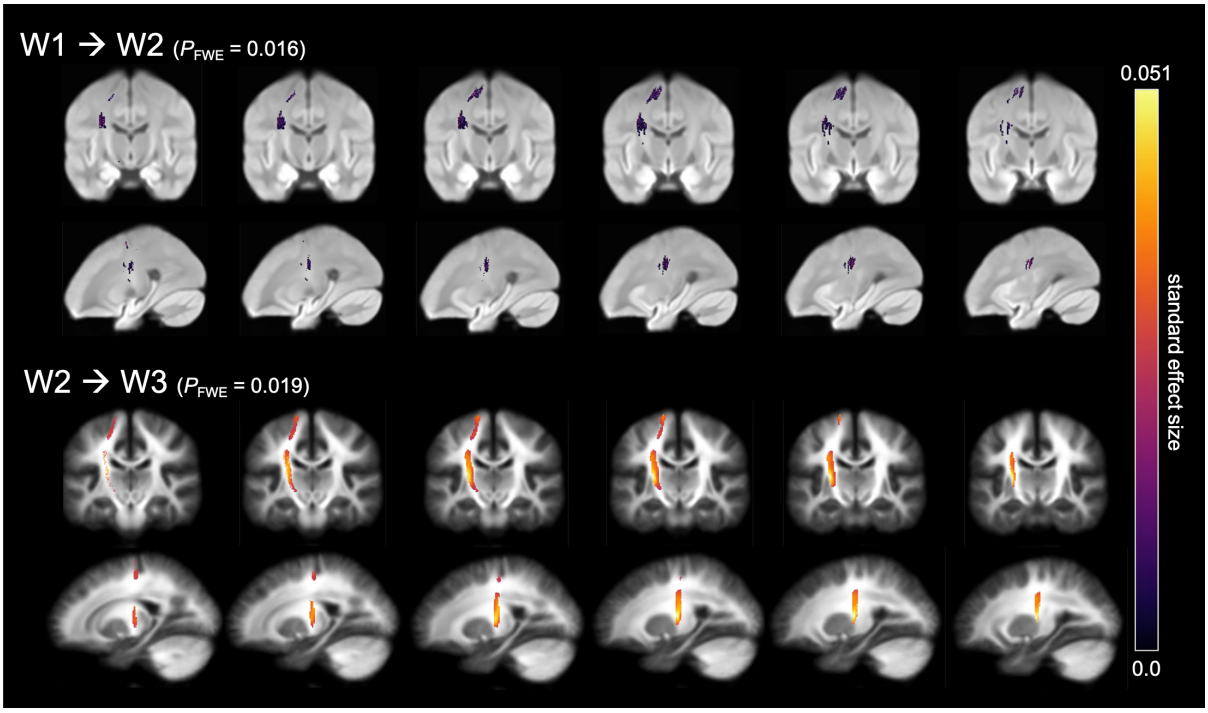

Figure 4. Symptom change may precede lCST WM microstructure plasticity. **Top**: Improvement of HI score is associated with more follow-up fiber density (FD). **Bottom**: Improvement of combined score is associated with more follow-up fiber cross-section (FC). Streamline segments have been cropped from the template tractogram to include only streamline points that correspond to significant fixels for this tract (FWE-corrected *p*-value < 0.05). Significant streamlines are colored by the standard effect size of 'Δscore' on 'FD at W2' and 'log(FC) at W3' and displayed across coronal and sagittal slices of the study-specific white matter fiber orientation distribution templates.

8. Supplementary material

Search space	Dependent variable Fixel-based metric	Contrast	Independent variable Symptom score change (Δ = W2 – W1)	t_{max}	Std. effect	P_{FWE}
Left corticospinal tract	**Fiber density**	(+)	combined	0.858	0.011	0.165
		(–)	combined	0.917	0.012	0.104
		(+)	HI	0.949	0.038	0.225
		(–)	**HI**	**1.092**	**0.044**	**0.016***
	Fiber cross-section	(+)	combined	1.846	0.023	0.591
		(–)	combined	2.264	0.029	0.343
		(+)	HI	1.909	0.076	0.627
		(–)	HI	2.189	0.087	0.536
	Fiber density and cross-section	(+)	combined	0.969	0.012	0.280
		(–)	combined	0.828	0.011	0.439
		(+)	HI	1.051	0.042	0.294
		(–)	HI	0.948	0.038	0.083
Left superior longitudinal fasciculus I	Fiber density	(+)	combined	2.238	0.028	0.496
		(–)	combined	1.300	0.017	0.285
	Fiber cross-section	(+)	combined	2.627	0.033	0.252
		(–)	combined	1.759	0.023	0.474
	Fiber density and cross-section	(+)	combined	2.483	0.032	0.112
		(–)	combined	1.351	0.017	0.493
Left superior longitudinal fasciculus II	Fiber density	(+)	combined	1.547	0.020	0.462
		(–)	combined	1.462	0.019	0.682
	Fiber cross-section	(+)	combined	1.987	0.025	0.319
		(–)	combined	1.372	0.017	0.625
	Fiber density and cross-section	(+)	combined	1.666	0.021	0.245
		(–)	combined	1.304	0.017	0.777
Left superior longitudinal fasciculus III	Fiber density	(+)	combined	1.464	0.019	0.762
		(–)	combined	1.290	0.016	0.225
	Fiber cross-section	(+)	combined	1.331	0.017	0.709
		(–)	combined	3.121	0.040	0.051
	Fiber density and cross-section	(+)	combined	1.766	0.022	0.863
		(–)	combined	1.239	0.016	0.407

Table S1. Results from the tract-specific fixel-based analysis of W1 to W2 in lCST and lSLF.

Search space	Dependent variable Fixel-based metric	Contrast	Independent variable Symptom score change (Δ = W3 – W2)	t_{max}	Std. effect	P_{FWE}
Left corticospinal tract	Fiber density	(+)	combined	1.288	0.017	0.612
		(–)	combined	1.687	0.023	0.156
		(+)	HI	1.344	0.036	0.636
		(–)	HI	1.761	0.048	0.200
	Fiber cross-section	(+)	combined	0.916	0.012	0.896
		(–)	**combined**	**3.775**	**0.051**	**0.019***
		(+)	HI	0.911	0.025	0.960
		(–)	HI	3.575	0.097	0.058
	Fiber density and cross-section	(+)	combined	1.122	0.015	0.821
		(–)	combined	1.776	0.024	0.093
		(+)	HI	1.190	0.032	0.908
		(–)	HI	1.800	0.049	0.141
Left superior longitudinal fasciculus I	Fiber density	(+)	combined	1.632	0.022	0.428
		(–)	combined	1.519	0.020	0.956
	Fiber cross-section	(+)	combined	0.740	0.010	0.963
		(–)	combined	2.863	0.038	0.169
	Fiber density and cross-section	(+)	combined	1.727	0.023	0.413
		(–)	combined	1.823	0.024	0.715
Left superior longitudinal fasciculus II	Fiber density	(+)	combined	1.264	0.017	0.837
		(–)	combined	1.247	0.017	0.882
	Fiber cross-section	(+)	combined	0.882	0.012	0.978
		(–)	combined	2.863	0.038	0.259
	Fiber density and cross-section	(+)	combined	1.340	0.018	0.587
		(–)	combined	1.305	0.017	0.730
Left superior longitudinal fasciculus III	Fiber density	(+)	combined	2.062	0.028	0.067
		(–)	combined	1.572	0.021	0.340
	Fiber cross-section	(+)	combined	1.450	0.019	0.778
		(–)	combined	2.210	0.030	0.437
	Fiber density and cross-section	(+)	combined	1.928	0.026	0.191
		(–)	combined	1.734	0.023	0.516

Table S2. Results from the tract-specific fixel-based analysis of W2 to W3 in lCST and lSLF.

Search space	Dependent variable Fixel-based metric	Contrast	Independent variable CAARS score (W3)	t_{max}	Std. effect	P_{FWE}
Whole brain	Fiber density	(+)	combined	3.011	0.045	0.562
		(–)	combined	2.739	0.041	0.858
		(+)	HI	3.195	0.082	0.501
		(–)	HI	2.640	0.068	0.835
		(+)	IA	3.120	0.095	0.053
		(–)	IA	2.862	0.087	0.757
	Fiber cross-section	(+)	combined	2.775	0.042	0.835
		(–)	combined	4.446	0.067	0.261
		(+)	HI	2.172	0.056	0.892
		(–)	HI	4.218	0.109	0.173
		(+)	IA	3.721	0.113	0.700
		(–)	IA	3.749	0.114	0.362
	Fiber density and cross-section	(+)	combined	3.027	0.046	0.678
		(–)	combined	2.847	0.043	0.778
		(+)	HI	3.111	0.080	0.864
		(–)	HI	2.774	0.071	0.827
		(+)	IA	3.166	0.096	0.119
		(–)	IA	2.840	0.086	0.716
Right cingulum bundle	Fiber density	(+)	HI	0.047	1.841	0.410
		(–)	HI	0.043	1.678	0.315
	Fiber cross-section	(+)	HI	0.763	0.020	0.980
		(–)	HI	3.806	0.098	0.063
	Fiber density and cross-section	(+)	HI	1.653	0.043	0.688
		(–)	HI	1.764	0.045	0.172

Table S3. Results from the cross-sectional global and tract-specific fixel-based analysis of W3 only. All general linear models showed no significant effects of CAARS score on any fixel-based metric.

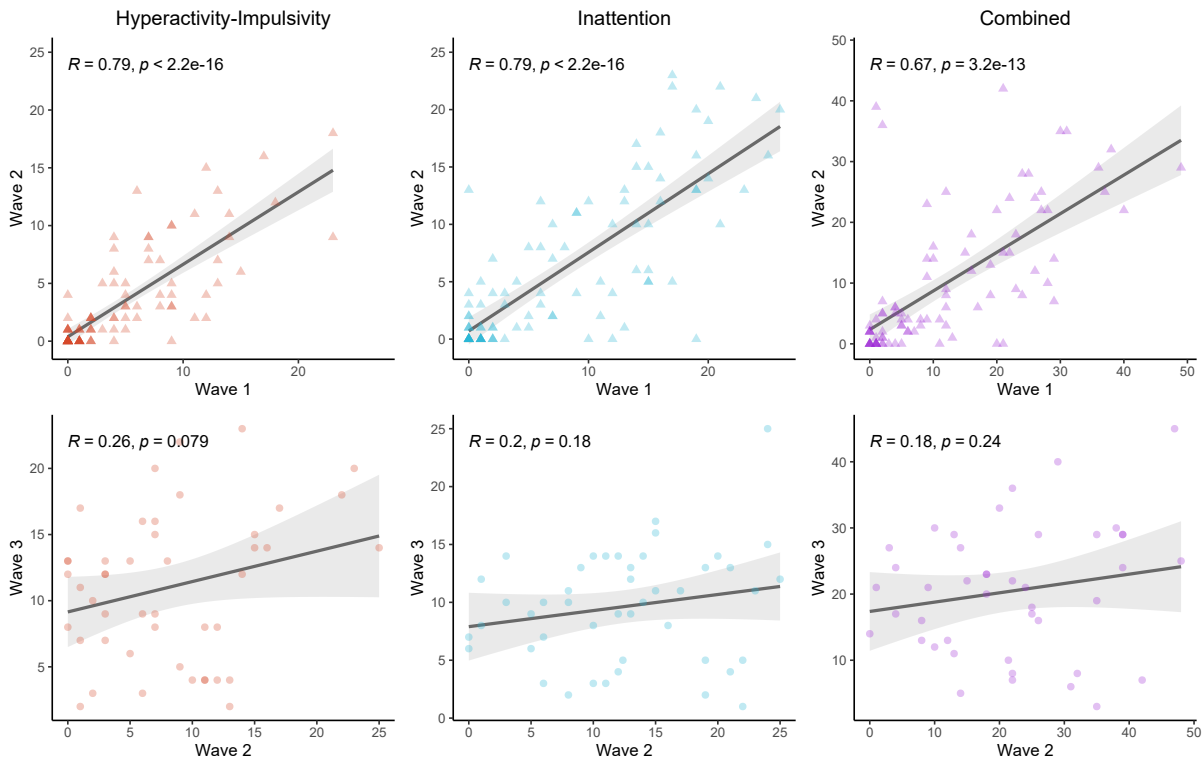

Figure S1. Correlation scatterplots of baseline against follow-up scores colored by dimension with 95% confidence intervals and Pearson correlation coefficients reported for each. **Top row:** Wave 1 associations with Wave 2 scores. **Bottom row:** Wave 2 associations with Wave 3 scores. **Left column**: Hyperactivity-impulsivity dimension scores. **Center column:** Inattention dimension scores. **Right column:** Combined scores (calculated as the sum of hyperactivity-impulsivity and inattention). Darker colored points indicate individuals with overlapping score data.

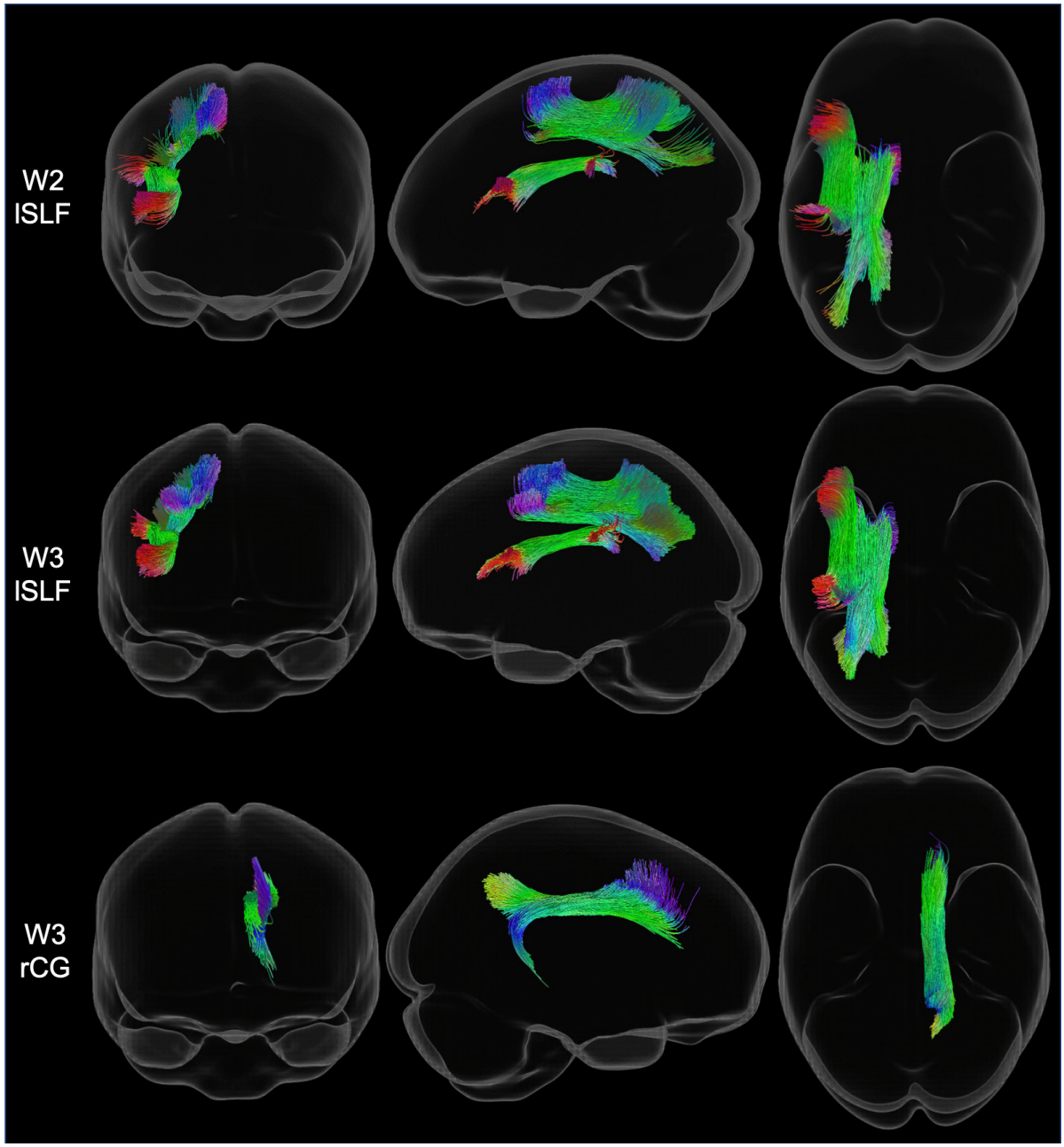

Figure S2. Tract-specific region-of-interest masks of the left superior longitudinal fasciculus (lSLF) at Wave 2 and Wave 3, and the right cingulum bundle (rCG) at Wave 3 colored by direction (red: left-right, green: anterior-posterior, blue: inferior-superior). Coronal (left column), sagittal (middle column), and axial (right column) views of the tract reconstructions from TractSeg (applied to the fiber orientation distribution templates) and displayed in glass brains for visualization.

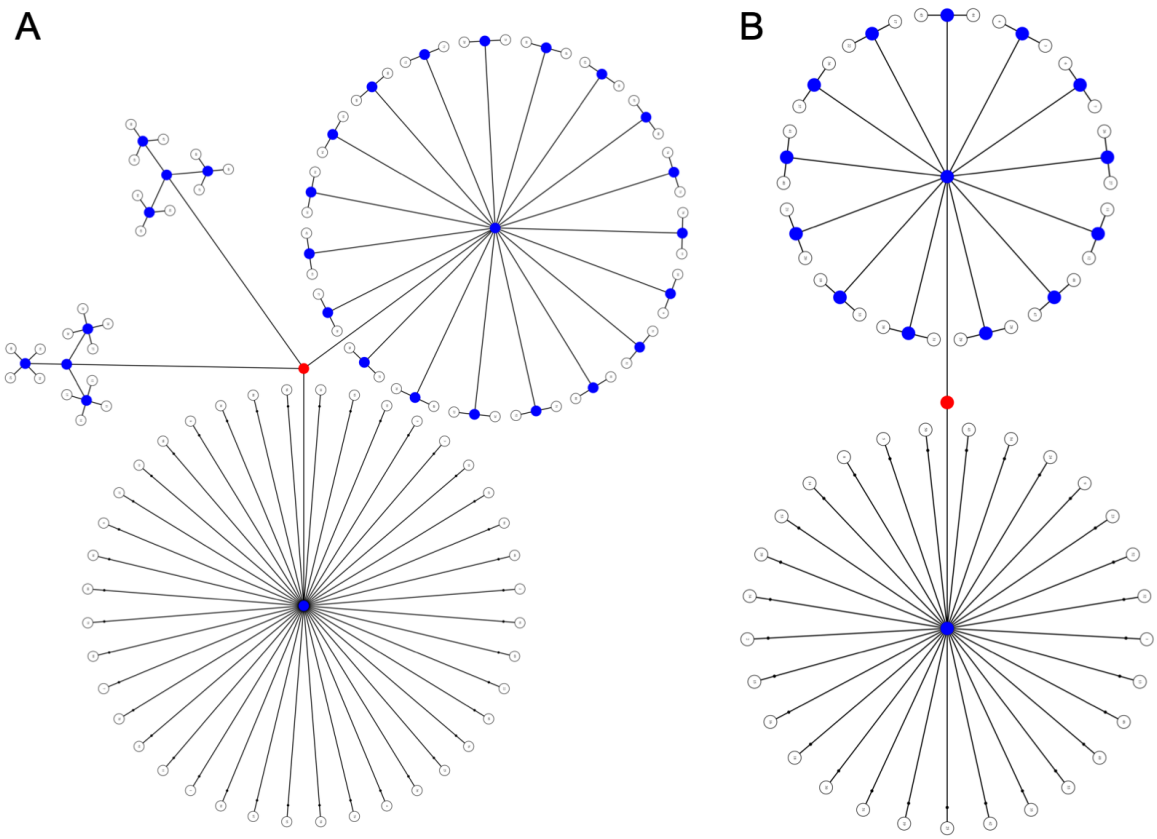

Figure S3. Visualization of the exchangeability block structure of (A) Wave 2 and (B) Wave 3, each represented in permutation trees. Each white dot is one individual and each group of white dots represents families of a specific size. At each permutation, branches beginning at blue dots can be permuted, while those beginning in red dots cannot. These permutation sets were generated with FSL PALM and used in connectivity-based fixel enhancement analysis to control for related siblings in each sample.

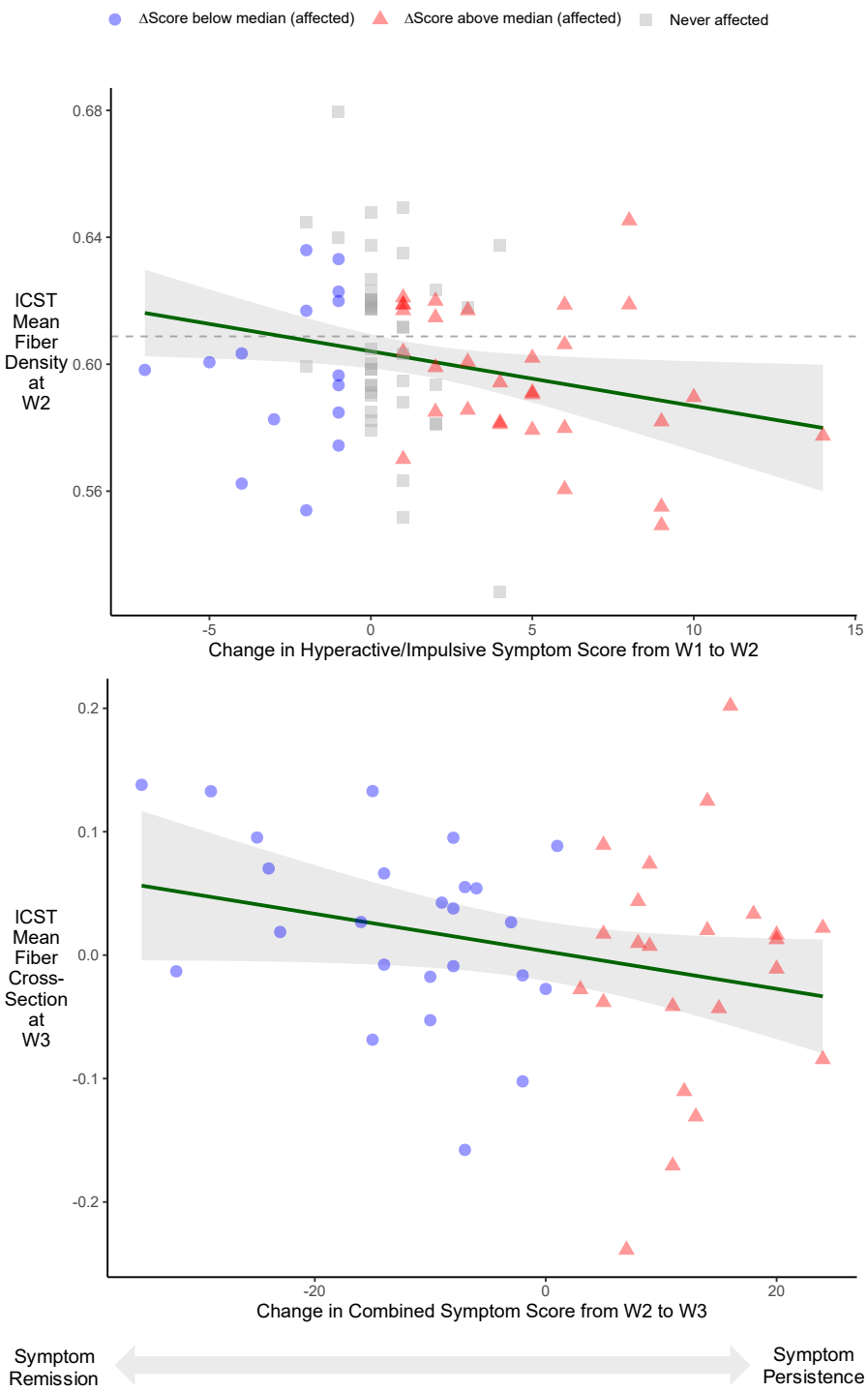

Figure S4. Graphs illustrating the relationship between left corticospinal tract (**Top**) mean fiber density values and change in hyperactive/impulsive symptom score from Wave 1 to Wave 2, and (**Bottom**) mean fiber cross-section values and change in combined symptom score from Wave 2 to Wave 3. Both plots include solid green regression lines with 95% confidence intervals. For reference, the mean fiber density at Wave 2 for unaffected participants is marked by a dashed gray line. Change in symptom score was calculated as: $\Delta Score = score_{follow\text{-}up} - score_{baseline}$. Participants are shown as individual points, differentiated according to whether their change in symptom score was above or below the median change in score.

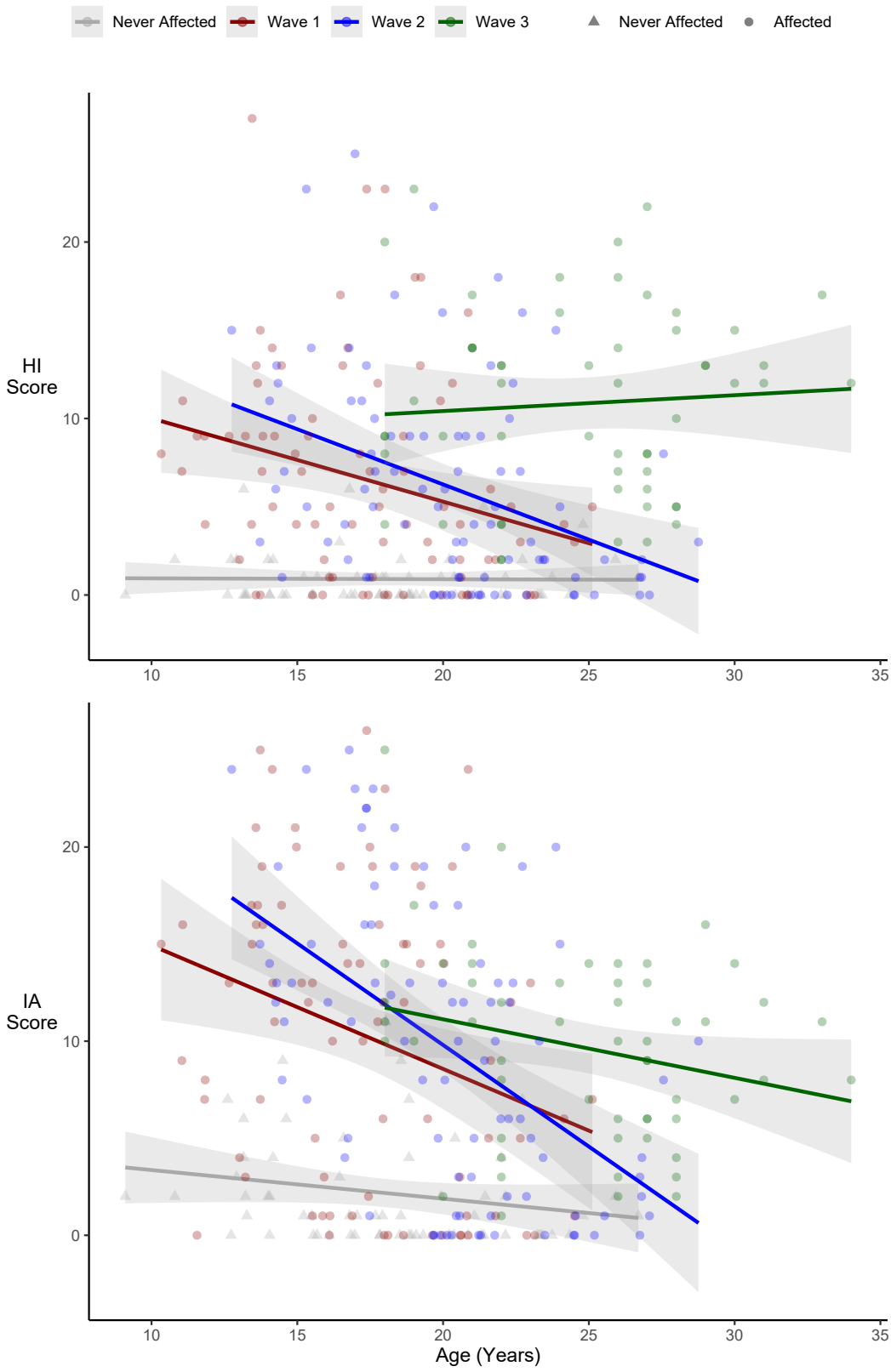

Figure S5. Participants from all waves pooled together (never affected are in gray triangles, and affected are circles colored according to wave in red [W1], blue [W2], and green [W3]) and plotted by age against symptom score in dimensions HI (**Top**) and IA (**Bottom**). Wave 3 included less participants and only those with a history of ADHD; scores were obtained with CAARS instead of CPRS.