












RESEARCH ARTICLE OPEN ACCESS

Pattern of Deep Grey Matter Undersizing Boosts MRI-Based Diagnostic Classifiers in Fetal Alcohol Spectrum Disorders

Eliot Kerdreux^{1,2}  | Justine Fraize^{1,2}  | Alexandra Ntorkou³  | Pauline Garzón^{1,2}  | Richard Delorme⁴  |
Monique Elmaleh-Berges^{2,3}  | Edouard Duchesnay⁵  | Lucie Hertz-Pannier^{1,2}  | Yann Leprince^{1,2}  |
Jean-François Mangin⁵  | David Germanaud^{1,2,6} 

¹Université Paris-Saclay, CEA, Joliot Institute, NeuroSpin, UNIACT, Gif-sur-Yvette, France | ²Université Paris Cité, Inserm, NeuroDiderot, inDEV Team, Paris, France | ³Department of Pediatric Radiology, Centre of Excellence InovAND, AP-HP, Robert Debré Hospital, Paris, France | ⁴Department of Child and Adolescent Psychiatry, Centre of Excellence InovAND, AP-HP, Robert Debré Hospital, Paris, France | ⁵Université Paris-Saclay, CEA, Joliot Institute, NeuroSpin, BAOBAB, Gif-sur-Yvette, France | ⁶Department of Genetics, Centre of Excellence InovAND, AP-HP, Robert Debré Hospital, Paris, France

Correspondence: Eliot Kerdreux (eliot.kerdreux@inserm.fr) | David Germanaud (david.germanaud@cea.fr)

Received: 14 January 2025 | **Revised:** 25 April 2025 | **Accepted:** 4 May 2025

Funding: This work was supported by Institut pour la Recherche en Santé Publique, IRESP-19-ADDICTIONS-08; Agence Nationale de la Recherche, ANR-19-CE17-0028-01, ANR-23-IAHU-0010.

Keywords: deep grey matter | diagnostic classifier | fetal alcohol spectrum disorders | magnetic resonance imaging | neuroimaging biomarkers | normative analysis | supervised learning

ABSTRACT

In fetal alcohol spectrum disorders (FASD), brain growth deficiency is a hallmark of subjects with both fetal alcohol syndrome (FAS) and nonsyndromic FASD (NS-FASD, that is, those without specific diagnostic features). Although previous studies have suggested that the deep grey matter is heterogeneously affected at the group level, it has not yet been established within proper scaling modeling, nor has it been given a place in the FASD diagnostic criteria where neuroanatomical features still contribute almost nothing to diagnostic specificity. We segmented a 1.5T T1-weighted brain MRI dataset of 90 monocentric FASD patients (53 FAS, 37 NS-FASD) and 95 typically developing controls (ages 6–20), using volBrain-vol2Brain as reference, and both Freesurfer-SAMSEG and FSL-FIRST to estimate result robustness. The segmentation resulted in seven anatomical volumes: total brain (TBV), total deep grey matter, caudate, putamen, globus pallidus, thalamus, and accumbens. After adjusting for confounds, we fitted the scaling relationship between deep grey matter nuclei volumes (V_i) and TBV ($V_i = b \times TBV^a$) and evaluated the effect of FAS on scaling. We then estimated the volumetric deviation from typical scaling (\sqrt{DTS}) for each deep grey nucleus volume in the FAS sample. Finally, we tested the improvement of FAS versus control classifiers based on total deep grey matter \sqrt{DTS} or total brain deviation from typical volume, by adding the five nuclear \sqrt{DTS} , both in terms of performance and generalizability to NS-FASD. Scaling was significantly different between the FAS and control groups for all deep grey matter nuclei ($p < 0.05$). We confirmed the undersizing of total deep grey matter in FAS ($\sqrt{DTS} = -6\%$) and identified a pattern of volumetric undersizing, most pronounced in the caudate (-13%) and globus pallidus (-11%), less so in the thalamus (-4%) and putamen (-2%) and sparing the accumbens (0%). These findings were consistent across segmentation tools, despite variations in magnitude. The pattern-based classifier was more efficient than the one based on total deep grey matter alone ($p < 0.001$) and identified 32.4% of the NS-FASD as having a FAS-like deep grey matter phenotype, compared to 18.9% with the classifier based on total deep grey matter alone ($p = 0.113$). Added to a classifier based on TBV only, the pattern improved the performance ($p = 0.033$) of the model and increased identification of NS-FASD with a FAS-like neuroanatomical phenotype from 37.8% to 62.2% ($p = 0.002$). This study details the volumetric undersizing

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *Human Brain Mapping* published by Wiley Periodicals LLC.

of deep grey matter in a large series of FASD patients. It reveals a differential pattern of vulnerability to prenatal alcohol exposure partially convergent across automatic segmentation tools. It also strongly suggests that this pattern of volumetric undersizing in the deep grey matter may contribute to a neuroanatomical signature of FAS that is usable to improve the probabilistic diagnosis of NS-FASD by means of MRI-based diagnostic classifiers.

1 | Introduction

Alcohol consumption during pregnancy remains a leading cause of neurodevelopmental disorders (Flak et al. 2014; Tsang et al. 2016). The diagnosis of fetal alcohol spectrum disorders (FASD) encompasses the diverse clinical consequences resulting from prenatal alcohol exposure (Riley and McGee 2005). International diagnostic guidelines, such as Astley and Clarren (2000), Cook et al. (2016), and Hoyme et al. (2016), rely on clinical criteria to differentiate fetal alcohol syndrome (FAS)—a positive diagnosis characterized by specific physical features—from the non-specific, non-syndromic forms of FASD (NS-FASD), where these physical traits are either absent or incomplete and whose diagnosis remains probabilistic. Establishing the likelihood of the causal link between neurodevelopmental disorders and prenatal alcohol exposure would greatly benefit from additional specific criteria, particularly in these non-syndromic cases.

Cerebral growth deficiency is a well-recognized consequence of prenatal alcohol exposure (Donald et al. 2015), microcephaly being the only neuroanatomical feature specified in existing guidelines, in which neuroradiological anomalies contribute almost nothing to the syndromic diagnosis. However, numerous studies have reported focal volumetric abnormalities, particularly in the corpus callosum and cerebellum but also in the deep grey matter (Fraize, Convert, et al. 2023; Fraize, Fischer, et al. 2023; Inkelis et al. 2020; Sullivan et al. 2020). The deep grey matter structures refer to subcortical nuclei, notably including the basal ganglia (caudate, putamen, globus pallidus, accumbens) and the thalamus. Indeed, from early investigations involving small FAS patient and control samples (Archibald et al. 2001; Mattson et al. 1996), to subsequent larger-scale studies covering the entire FASD spectrum (Astley, Aylward, et al. 2009; Nardelli et al. 2011), a volume deficit in deep grey matter has been consistently reported in proportion to the whole brain volume, at least for the caudate. However, variations have emerged among recent studies depending on the specific nuclei significantly affected: while the globus pallidus and the caudate appear to be similarly affected (Nardelli et al. 2011), reports are heterogeneous for the putamen and the thalamus (Gautam et al. 2015; Inkelis et al. 2020; Roussotte et al. 2012; Treit et al. 2017), and only two studies have reported no significant effect on the accumbens (Archibald et al. 2001; Biffen et al. 2018).

Several factors likely introduce variability, or even intrinsic inaccuracy, between volumetric studies of deep grey matter in the fetal alcohol context, that can weaken or bias their findings. First, studies may consider rather different populations, ranging from patients with robust FAS diagnoses to less consensual full fetal alcohol spectrum or non-further-assessed prenatal alcohol exposure cases (Gautam et al. 2015; Inkelis et al. 2020; Roussotte et al. 2012;

Treit et al. 2017). The likelihood of a causal link to alcohol, as well as the mean level of PAE, differs between these groups, leading to a risk of confounding etiological factors, the dilution of disease-specific neuroanatomical effects, which are expected to correlate with severity (Astley, Olson, et al. 2009), and ultimately a loss of both robustness and sensitivity in the detection of anomalies. This is likely a major source of inconsistency. However, the hypothesis of a “continuum” in the neuroanatomical phenotype makes sense (Astley, Olson, et al. 2009) and supports the interest of certain group-stratified analyses, such as learning specific features from a FAS group and testing the generalizability in a NS-FASD one. Second, studies typically control for total brain volume restriction by assuming a linear relationship with the volume of interest (Boateng et al. 2023; Gautam et al. 2015; Nardelli et al. 2011; Roussotte et al. 2012), if not simply using a proportion (Archibald et al. 2001; Astley, Olson, et al. 2009). However, even in the general population, deep grey matter volumes vary with brain size in a non-proportional nonlinear, negative allometric relationship, meaning that smaller brains display proportionally larger deep grey matter structures (de Jong et al. 2017; Liu et al. 2014; Williams et al. 2021). Failing to address such allometric scaling has been proven to increase the risk of over- or underestimating volume differences, particularly in nonnormocephalic patients (Germanaud et al. 2014; Toro et al. 2009). Finally, studies employ a diversity of brain segmentation methods and versions. Differences in volume estimates are well documented across methods or even between versions of the same tool (Biffen et al. 2020; Haddad et al. 2023; Sankar et al. 2017). The issue does not lie in the methodological bias per se, but rather in the fact that volumetric differences between patients with FASD and controls may vary significantly depending on the segmentation tool (Biffen et al. 2020) to the point that studies with relatively similar recruitment and adjustments for total brain volume may report quite different patterns of deep grey matter impairment (Chen et al. 2012; Boateng et al. 2023). Surprisingly, few studies in FASD have attempted to assess the reliability of their results across multiple segmentation tools (Sullivan et al. 2020), at least by incorporating a robustness analysis step.

Until now, no study has explored how measures of deep grey matter impairment could contribute to the diagnostic process in FASD. Little and Beaulieu (2020) recently introduced a multivariate classification model based on 87 cerebral regions trained and tested on two mixed sets of FAS, NS-FASD, prenatal alcohol exposure cases without confirmed FASD diagnosis, and typically developing controls. Interestingly, the globus pallidus and caudate were among the regions that contributed the most to the classification performance. However, with nonspecific and undiagnosed forms included in the training set alongside FAS cases, the study was not designed to identify a combination of anomalies likely to provide additional diagnostic specificity when needed. In this regard, our team recently proposed using classifiers trained to differentiate patients with FAS from typically

Summary

- Severe and heterogeneous undersizing of deep grey matter in FASD, accounting for smaller brain size.
- This pattern provides additional diagnostic value beyond global deep grey matter undersizing.
- Incorporating these markers into a high-specificity diagnostic classifier based on microcephaly improves identification of NS-FASD patients with a FAS-like neuroanatomical signature.

developing controls based on the volume deviation from typical allometric scaling of seven cerebellar regions, to identify a significant subgroup of patients with NS-FASD (15%) exhibiting a FAS-like gradient of cerebellar undersizing, and for whom a causal link to alcohol could therefore be considered even more likely (Fraize, Fischer, et al. 2023).

In this context, the main objective of the present study was twofold: (1) to characterize a robust pattern of over- or undersizing of deep grey matter structures in patients with FAS, and (2) to establish whether this pattern provided additional diagnostic information beyond the overall reduction in brain volume(s) both in patients with FAS and NS-FASD. First, we expected that in our large patient sample with fully diagnosed FASD, allometry-sensitive scaling models would reveal disease-related perturbations of the relationships observed in typically developing controls between deep grey matter volumes and total brain volume, and that the deviation to typical scaling would be different in magnitude across nuclei, resulting in a significant pattern of deep grey matter over- or undersizing in the FAS group. We assessed the robustness of this pattern using two alternative segmentation tools. Second, we expected that this pattern would improve the performance of a diagnostic classifier trained to distinguish with high specificity patients with FAS from typically developing controls, beyond what could be achieved using total brain volume and global deep grey matter sizing alone. Applied in the NS-FASD group, we expected that this boosted classifier would increase the number of patients identified as exhibiting a neuroanatomical phenotype consistent with FAS.

2 | Material and Methods

2.1 | Participants

Participants in this study were 90 patients diagnosed with FASD, aged from 6 to 20 years ($M=11.4$, $SD=3.6$, sex ratio=1.4), retrospectively recruited from a clinical series of patients attending the NDD-dedicated child neurology consultation at Robert-Debré University Hospital (Paris) between 2014 and 2020. Typically developing (TD) controls were 95 subjects, from 6 to 20 years old ($M=12.2$, $SD=3.4$, sex ratio=1.0), with no report of prenatal alcohol exposure, developmental or school delay, nor family history of neurological or psychiatric condition. A subgroup of 40 TD subjects ($M=12.7$, $SD=4.1$, sex ratio=0.8) was matched to the FASD group for the acquisition site (MRI scanner and sequence). The other 55

TD subjects ($M=11.8$, $SD=2.8$, sex ratio=1.1) were part of a previously published study (Bouyeure et al. 2018). The diagnostic procedure was based on the two main FASD guidelines and included a differential diagnosis workup based on systematic brain MRI, biological and genetic tests. Exclusion criteria were exposure to another embryo-fetotoxic agent, very or extremely preterm birth (i.e., < 32 gestational weeks), a concomitant genetic anomaly that could explain the NDD phenotype, and explicit refusal to participate in the study. We divided the FASD sample into FAS and non-syndromic (NS-FASD) subgroups and evaluated the concordance between group assignments using the 4-digit diagnostic code (Astley and Clarren 2000), and the revised guidelines of the Institute of Medicine (Hoyme et al. 2016). This FASD series and the diagnostic procedure have already been described in previous studies (Fraize, Convert, et al. 2023; Fraize et al. 2022; Fraize, Fischer, et al. 2023; Fraize, Leprince, et al. 2023).

2.2 | MRI Acquisitions

The patients with FASD and scanner-matched controls were scanned on a 1.5T Ingenia scanner (Philips Healthcare, Amsterdam, the Netherlands) with a 3DT1 FFE-TFE sequence (1 mm isotropic; TR=8.2 ms; TE=3.8 ms; TI=0.8 s; Flip=8°; TA=5.77 min; SENSE=2). The remaining controls ($n=55$) were scanned on a 3T Trio scanner (Siemens Healthineers, Oxford, UK) with a 3DT1 Siemens MPRAGE sequence (1 mm isotropic; TR=2.3 s; TE=3 ms; TI=0.9 s; Flip=9°; TA=7.75 min; GRAPPA 2).

2.3 | MRI Processing and Segmentation

The whole brain was segmented using volBrain as in previous studies (Fraize et al. 2022; Fraize, Fischer, et al. 2023; Fraize, Leprince, et al. 2023). The subcortical regions of interest (caudate nucleus, putamen, globus pallidus, thalamus, and accumbens) were segmented using volBrain-vol2Brain version 1.0 (Manjón et al. 2022), whose atlas dataset includes more age diversity than volBrain. We considered this segmentation based on an optimized multi-atlas label fusion scheme as our reference as it consistently outperformed other comparable nuclear segmentation methods (Akudjedu et al. 2018; Hannoun et al. 2019; Manjón and Coupé 2016; Næss-Schmidt et al. 2016). However, to compare the robustness of the results across segmentation methods, we also segmented the subcortical regions of interest using two other recent automatic segmentations, Freesurfer-SAMSEG version 7.2.0 (Puonti et al. 2016) and FSL-FIRST version 6.0.5 (Patenaude et al. 2011) that rely on different segmentation strategies, respectively a mesh-based computational atlas combined with a Gaussian appearance model designed to be robust against variability induced by acquisition platform changes, and a Bayesian framework of probabilistic relationships between shape and intensity. Since we did not have a priori hypotheses regarding an interaction between FASD and cerebral lateralization, we pooled left and right cerebral structures for analysis.

We systematically reviewed the quality of brain segmentations from volBrain-vol2Brain, Freesurfer-SAMSEG, and FSL-FIRST by visually inspecting them on T1-weighted images using ITK-SNAP (Yushkevich et al. 2006). For each region of interest, we

identified outliers on group volume distributions, left-right asymmetry, and volume as a function of total brain volume. These outliers benefited from a double quality check. We detected clear inaccurate segmentations in five subjects with vol2Brain (2 thalami, 2 putamen, 1 caudate), four with FSL-FIRST (2 putamen, 2 caudate), and one with Freesurfer-SAMSEG (thalamus). These inaccuracies were manually corrected using ITK-SNAP.

2.4 | Statistical Analysis

All statistical analyses were performed using R statistical Software (v4.1.2; R Core Team 2021) along with the scikit-learn (v1.3.2; Pedregosa et al. 2011) and scipy (v1.11.4; Virtanen et al. 2020) Python packages. We used the Benjamini and Hochberg false discovery rate (FDR) method (Benjamini and Hochberg 1995) for p value correction to account for the increased risk of type I error in multiple comparisons. All p values were FDR-adjusted, except for the post hoc pairwise mean comparisons of adjusted volumes between groups, which were adjusted for family-wise error rate (FWER) using the Bonferroni correction.

2.4.1 | Adjusting for Confounders or Variables of Noninterest

We adjusted the data for the effect of the acquisition site, but also for age and sex in the perspective of normative scaling analysis. Indeed, residual growth was expected to be very small, or nonsignificant, in our sample. Besides, as in previous studies (Fraize et al. 2022; Fraize, Fischer, et al. 2023; Fraize, Leprince, et al. 2023), we favored statistical power over discrimination between males and females. As a robustness analysis of the latter two choices, we performed post hoc analysis on site-only adjusted data. Practically, we iteratively adjusted the data for the site and then for sex differences using the NeuroCombat R-package (Fortin et al. 2018), a batch effect harmonizer based on an empirical Bayes framework, integrating age and diagnosis into the model as covariates to be spared by corrections. Subsequently, we addressed age-related effects on the few volumes that showed significant ones (accumbens for volBrain-vol2Brain, thalamus for FreeSurfer-SAMSEG, and globus pallidus for FSL-FIRST) using adjusted residuals harmonization (Fortin et al. 2018). This comprehensive approach involved considering all covariates in the model (Equation 1) and subtracting the age effect from this model from the volumes of interest (V_i) (Equation 2).

$$V_i = b_0 + b_{age} \times age + b_{diagnostic} \times diagnostic + b_{site} \times site + b_{sex} \times sex \quad (1)$$

$$V_i^{adj} = V_i - b_{Age} \times age \quad (2)$$

After each adjustment, we ensured that non-interest effects disappeared while preserving the diagnostic effect (data not shown).

2.4.2 | Assessing the Effect of FAS on Scaling

We performed scaling analyses examining the relationship between deep grey matter volumes (V_i) and brain size (total brain

volume, TBV) with a nonlinear power law model that is sensitive to allometric effects that is variations of proportion with size (de Jong et al. 2017; Germanaud et al. 2012; Liu et al. 2014). Practically, we fitted the equation $V_i = b \times TBV^a$ using nonlinear least square and tested whether the scaling coefficients “ a ” in the FAS and control groups (a_{FAS} and a_{TD} , respectively) were different from isometry ($a = 1$). Then, we assessed the influence of bearing a FAS on allometric laws after log-linearization by comparing two nested hierarchical models, with and without a diagnosis covariate (Equations 3 and 4), using a likelihood ratio test and computing the differences of adjusted determination coefficients (ΔR^2), Akaike information criterion (AIC), and Bayesian information criterion (BIC).

$$\log_{10}(V_i) = a \times \log_{10}(TBV) + \log_{10}(b) \quad (3)$$

$$\log_{10}(V_i) = a \times \log_{10}(TBV) + \log_{10}(b) + a_{FAS} \times FAS \\ \times \log_{10}(TBV) + FAS \times \log_{10}(b_{FAS}) \quad (4)$$

2.4.3 | Quantifying the Volumetric Loss in FASD Compared to Controls

For each subject and region of interest, we determined the individual volume deviation from the typical scaling fit (\sqrt{DTS}) which corresponds to the distance from the typical scaling prediction, relative to the mean volume of controls (Fraize, Fischer, et al. 2023). Subsequently, we computed the means of these \sqrt{DTS} values along with their corresponding 95% confidence intervals. We compared the mean \sqrt{DTS} and distributions between patient subgroups (FAS and NS-FASD) and controls (whose \sqrt{DTS} means were around 0), estimating the variance of both the typical scaling fit in controls and the \sqrt{DTS} computation in the different groups with two embedded bootstrap resamplings: the first one in controls for the typical scaling fit ($n = 1000$) with a second embedded one ($n = 1000$) for the mean \sqrt{DTS} computation. p values were computed using confidence interval inversion (Thulin 2021). Finally, within each FASD subgroup, we conducted pairwise comparisons to evaluate whether the mean \sqrt{DTS} distributions varied across different regions of interest.

2.4.4 | Assessing the Reliability of the Results Across Segmentation Methods

After calculating the average correlation and DICE coefficients (Python package seg-metrics v1.1.3., Jia 2020) between the segmentation results, we performed scaling and scaling deviation analyses on the volumes extracted by the other two segmentation methods. Then, we tested the significance of the segmentation method's effect on the \sqrt{DTS} of each region of interest using bootstrap pairwise comparisons.

2.4.5 | Classifying Patients With FASD

We built two nested linear logistic models to predict diagnosis status (FAS vs. TD). The first model used the total deep grey matter \sqrt{DTS} as the sole predictor (unifactorial classifier), while the second

added the five nuclear $\sqrt[3]{\text{DTS}}$ (multifactorial classifier). We assessed the quality of the multifactorial classifier over the unifactorial one using the likelihood ratio test of goodness of fit and comparing the Akaike information criterion (AIC) and Bayesian information criterion (BIC) between models. Model performance was assessed on a training sample of patients with FAS and TD subjects, setting the threshold probability to achieve specificity above 95% and comparing the area under the receiver operating characteristic (ROC) curves (AUC) between models. We also performed a stratified 5-fold cross-validation of the two model performances and compared their mean performance criteria (AUC, specificity, and accuracy). Finally, we applied the two classifiers on our NS-FASD subgroup as a test sample and compared the proportions of patients with NS-FASD classified as harboring a FAS-like deep grey matter phenotype using the McNemar exact test (Varoquaux and Colliot 2023).

To determine whether adding deep grey matter information improved diagnostic classification based on total brain size, we first verified that there was no significant age effect on total brain volume and then computed individual deviation from typical brain volume ($\text{TBV}_{\text{dist}} = (\text{TBV}_i - \text{mean TBV}_{\text{TD}}) / \text{mean TBV}_{\text{TD}}$) and set a classifier based on it (model 3, Equation 5). We added predictors step by step, total deep grey matter $\sqrt[3]{\text{DTS}}$ (model 4, Equation 6) and then the five nuclear $\sqrt[3]{\text{DTS}}$ (model 5, Equation 7) to see if it improved the quality of the model and the classification performances. We checked for over-fitting risk using 5-fold cross-validation as for previous classifiers and then compared the improvement of quality and performance of the three nested models.

$$\text{logit}(p_{\text{FAS}}) = b_0 + b_1 \times \text{TBV}_{\text{dist}} \quad (5)$$

$$\text{logit}(p_{\text{FAS}}) = b_0 + b_1 \times \text{TBV}_{\text{dist}} + b_2 \times \text{total deep grey matter}_{\sqrt[3]{\text{DTS}}} \quad (6)$$

$$\begin{aligned} \text{logit}(p_{\text{FAS}}) = & b_0 + b_1 \times \text{TBV}_{\text{dist}} + b_2 \times \text{total deep grey matter}_{\sqrt[3]{\text{DTS}}} \\ & + b_3 \times \text{caudate}_{\sqrt[3]{\text{DTS}}} + b_4 \times \text{putamen}_{\sqrt[3]{\text{DTS}}} + b_5 \times \text{globus pallidus}_{\sqrt[3]{\text{DTS}}} \\ & + b_6 \times \text{thalamus}_{\sqrt[3]{\text{DTS}}} + b_7 \times \text{accumbens}_{\sqrt[3]{\text{DTS}}} \end{aligned} \quad (7)$$

3 | Results

The analysis included 185 participants (95 typically developing controls and 90 subjects with FASD). Table 1 contains demographics and clinicals for the two FASD subgroups (FAS and NS-FASD). Sex and age at MRI acquisition did not significantly differ between FASD and TD ($\chi^2_{\text{sex}}(1) = 1.29, p_{\text{sex}} = 0.26; t_{\text{age}}(183) = 1.63, p_{\text{age}} = 0.11$), nor between FASD subgroups ($\chi^2_{\text{sex}}(1) = 1.39, p_{\text{sex}} = 0.195, t_{\text{age}}(88) = -1.17, p_{\text{age}} = 0.24$) and between TD subgroups ($\chi^2_{\text{sex}}(1) = 0.29, p_{\text{sex}} = 0.59, t_{\text{age}}(93) = -1.13, p_{\text{age}} = 0.26$).

3.1 | Mean Volume Comparison Between Groups

FAS and NS-FASD groups showed lower raw and adjusted total brain volume (TBV), total deep grey matter volume, and nuclear volumes than controls. The volumes in the NS-FASD group were intermediate between controls and FAS groups except for the accumbens and the putamen (see Figure 1).

3.2 | Deep Grey Matter Scaling Analysis: Accounting for Brain Size Effect

3.2.1 | Reference Scaling Laws in Controls

In the TD group, the scaling of both total deep grey matter volume relative to the TBV (scaling exponent $a = 0.71, p < 0.001$) and the nuclei volumes exhibited negative allometry (all $a < 1$, all $p < 0.05$) (see Figure 2).

3.2.2 | Effects of FAS on Deep Grey Matter Scaling Laws

In the FAS group, neither the scaling of total deep grey matter volume nor that of nuclei volumes differed significantly from isometry ($a = 1$). The effect of FAS diagnosis as a covariate on the scaling exponents was significant for each nucleus ($p_{\text{DIAG}} < 0.05$) except for the putamen and accumbens ($p_{\text{DIAG}} > 0.05$) (see Figure 2).

3.3 | Normative Analysis: Undersized Volumes in FAS

In the FAS group, all the nuclei except the accumbens ($p = 0.95$) exhibited significantly smaller average volumes than expected for the total brain according to typical scaling. This undersizing was not uniform across the nuclei. A distinct severity pattern or gradient was observed: the caudate and globus pallidus showed the most negative deviation from typical scaling ($\sqrt[3]{\text{DTS}}$) values, followed by the thalamus and putamen with slightly less pronounced effects (Table 2). In the NS-FASD sample, a similar pattern of mean deep grey matter $\sqrt[3]{\text{DTS}}$ was observed, with intermediate undersizing.

Pairwise comparisons of mean $\sqrt[3]{\text{DTS}}$ between the nuclei were significant between the caudate and thalamus, putamen, and accumbens (all $p < 0.05$), as well as between the globus pallidus and thalamus, putamen, and accumbens (all $p < 0.05$). However, there were no significant mean differences between the caudate and globus pallidus, or between the thalamus, accumbens, and putamen (see Figure 4A). A similar pattern of pairwise mean comparisons was observed in the NS-FASD sample (see Table S1). The average pattern of $\sqrt[3]{\text{DTS}}$ observed in patients with FAS is illustrated in Figure 3A.

The deep grey matter pattern of undersizing was marginally affected by age and sex volume adjustments (see Table S2).

3.4 | Post Hoc Robustness Analysis: Impact of the Segmentation Tool

The mean Dice similarity coefficients between segmentations were all $\geq 84\%$ except for the accumbens (67%, 68%, and 75%). Similarly, the volumes of the nuclei were highly correlated between segmentation tools with all $r \geq 0.86$ except for the accumbens (0.69, 0.45, and 0.66) (see Tables S3 and S4).

The pairwise comparison of the deep grey matter $\sqrt[3]{\text{DTS}}$ between segmentation tools showed a trend towards more severe $\sqrt[3]{\text{DTS}}$

TABLE 1 | Demographic, clinical, radiological data of patients with fetal alcohol syndrome (FAS) and nonsyndromic fetal alcohol spectrum disorders (NS-FASD).

| | FAS (<i>n</i> = 53) | NS-FASD (<i>n</i> = 37) | Group comparison (<i>p</i>) |
|--|----------------------|--------------------------|-------------------------------|
| Sociodemographic assessment | | | |
| Sex: male, <i>n</i> (%) | 28 (52.8) | 25 (67.6) | 0.195 |
| Age at scan, mean in years (SD) | 10.99 (3.55) | 11.88 (3.55) | 0.245 |
| 4-digit diagnostic code based clinical assessment, <i>n</i> (%) | | | |
| (1) Prenatal alcohol exposure | | | |
| Confirmed, severe | 22 (41.5) | 16 (43.2) | 1.000 |
| Confirmed, moderate or unquantified | 25 (47.2) | 18 (48.6) | 1.000 |
| Not documented | 6 (11.3) | 3 (8.1) | 0.732 |
| No exposure | 0 (0.0) | 0 (0.0) | — |
| (2) FAS facial features | | | |
| Severe | 32 (60.4) | 2 (5.4) | < 0.001 |
| Moderate | 21 (39.6) | 1 (2.7) | < 0.001 |
| Mild | 0 (0.0) | 30 (81.1) | < 0.001 |
| None | 0 (0.0) | 4 (10.8) | 0.028 |
| (3) Growth deficiency | | | |
| Significant | 19 (35.8) | 3 (8.1) | 0.003 |
| Moderate | 11 (20.8) | 2 (5.4) | 0.06 |
| Mild | 9 (17.0) | 9 (24.3) | 0.43 |
| None | 14 (26.4) | 23 (62.2) | 0.001 |
| Brain anatomy | | | |
| (4) Severe structural CNS damage | 41 (77.4) | 17 (45.9) | 0.003 |
| (4) Microcephaly (≤ -2 SD) | 37 (69.8) | 13 (35.1) | 0.002 |
| Diagnostic agreement with revised Institute of Medicine Guidelines | 55 (61.1) | 35 (39.8) | |
| Cohen's Kappa coefficient | 0.815 | | < 0.001 |

Note: In bold, *p* values ≤ 0.05 .

Abbreviation: CNS, central nervous system.

with FSL-FIRST, followed by Freesurfer-SAMSEG and volBrain-vol2Brain. However, there were no significant differences between the latter two for the caudate and putamen, between Freesurfer-SAMSEG and FSL-FIRST for the accumbens, nor between volBrain-vol2Brain and FSL-FIRST for the globus pallidus (see Table S5 mean comparisons column and Table S6). The pattern of undersizing severity among the nuclei was ordinaly consistent between segmentation tools, producing the same qualitative ranking by order of severity (see colored comparison matrices in Figure 4). However, there were divergences in deciding which differences between the nuclei were statistically significant, showing at least a divergence in estimating the magnitude of the difference between nuclei. Nevertheless, the caudate was significantly more affected than all the other nuclei except the globus pallidus, consistently across segmentation tools. The globus pallidus was significantly more affected than the accumbens consistently across segmentation tools, more than the putamen with two segmentation tools

(volBrain-vol2Brain and Freesurfer-SAMSEG) and more than the thalami with one tool (volBrain-vol2Brain).

3.5 | Deep Grey Matter-Based Classifiers

3.5.1 | Classification Performance: Patients With FAS vs. Typically Developing Controls

The logistic model that included both total deep grey matter and the five nuclear $\sqrt{\text{DTS}}$ was significantly more effective in identifying diagnostic status (FAS vs. TD) than the model on total deep grey matter alone ($\chi^2(5) = 57.11$, $p < 0.001$, with lower AIC and BIC). AUCs (Figure 5E) derived from the two models' ROC curves showed that the multiple-predictor classifier outperformed the one based solely on total deep grey matter (AUC = 91.48% vs. 74.96%, $Z = -3.88$, $p < 0.001$), which was confirmed by five-fold

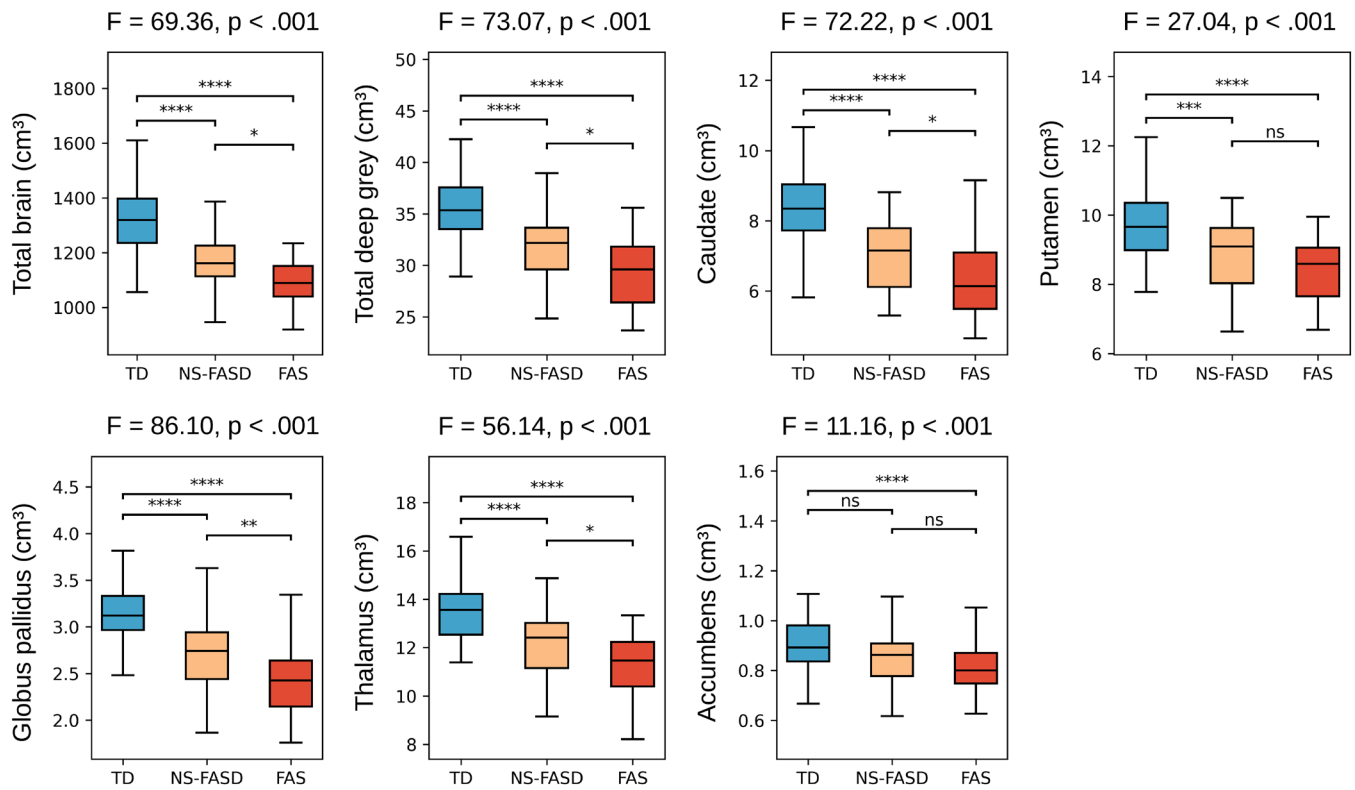


FIGURE 1 | Total brain and deep grey matter adjusted volume comparisons between typically developing (TD), fetal alcohol syndrome (FAS), and non-syndromic fetal alcohol spectrum disorders (NS-FASD) groups. One factor, 3-level Anovas on volumes and Bonferroni-adjusted p values for post hoc t tests (ns: Nonsignificant, $*p \leq 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$). Volumes were adjusted for site and sex effects using NeuroCombat (Fortin et al. 2018), and for age using linear residualization.

cross validation ($AUC_{CV} = 74.96\%$ vs. 91.17%). The estimated accuracy and sensitivity were also improved by the adjunction of the five nuclear \sqrt{DTS} (Table S7 and Figure 5A,B,E).

3.5.2 | Application to the NS-FASD Sample

The two classifiers trained on the FAS and TD samples and calibrated for high diagnostic specificity ($>95\%$) accurately identified respectively 49.1% and 71.7% of patients with FAS ($p = 0.002$) (Figure 5A). The adjunction of the five nuclear \sqrt{DTS} to the classifier led to a substantial 46.1% increase in the rate of correctly classified patients with FAS. When applied to the NS-FASD sample, the classifier relying solely on total deep grey matter identified only 18.9% ($n = 7$) of patients with NS-FASD as probable FAS. In contrast, the inclusion of the five nuclear \sqrt{DTS} increased the identification rate to 32.4% ($n = 12$), representing a 71.4% increase, although non-significant ($p = 0.113$) (Figure 5B).

3.5.3 | Comparing Quality and Performance of Classifiers Based on Deep Grey Matter Volume vs. Total Brain Volume

The decision threshold associated with 95% specificity of the classifier based solely on TBV distance (Model 3) corresponded to the 4th percentile of the controls TBV distance from typical volume distribution. Adding the total deep grey matter \sqrt{DTS} (Model 4) and then each nuclear \sqrt{DTS} (Model 5) significantly improved goodness of fit compared to a minimal classifier

relying solely on TBV (Model 3), $\chi^2_{M4-M3}(1) = 9.58$, $p = 0.002$, $\chi^2_{M5-M4}(5) = 22.88$, $p < 0.001$ with lower AIC and BIC. There was a significant improvement in patients with FAS correctly identified between Model 3 and Model 5 (69.8% vs. 83.0%, $p = 0.033$) associated with a significant difference in patients with NS-FASD identified as probable FAS (37.8% vs. 62.2%, $p = 0.002$) (see Figure 5C,D and Table S8).

4 | Discussion

Our study revealed a heterogeneous pattern of deep grey matter volumetric undersizing in a large sample of patients with FASD, accounting for the abnormally small brain size with an appropriate scaling model. Specifically, the caudate nucleus and globus pallidus were the most undersized, followed by intermediate undersizing in the thalamus and putamen, and minimal or negligible undersizing in the accumbens. This pattern order was consistently observed with two additional automatic segmentation tools, although the magnitude of the heterogeneity varied significantly between them. When incorporated into a diagnostic classifier trained to distinguish with high specificity between patients with FAS and TD subjects, nuclear predictors improved the classification performance compared to relying solely on the global deep grey matter predictor. Moreover, we demonstrated that the global deep grey matter and the nuclear predictors further enhanced the performance of a “reference” diagnostic classifier based on the total brain volume only. Applied to a sample of patients with NS-FASD, this approach

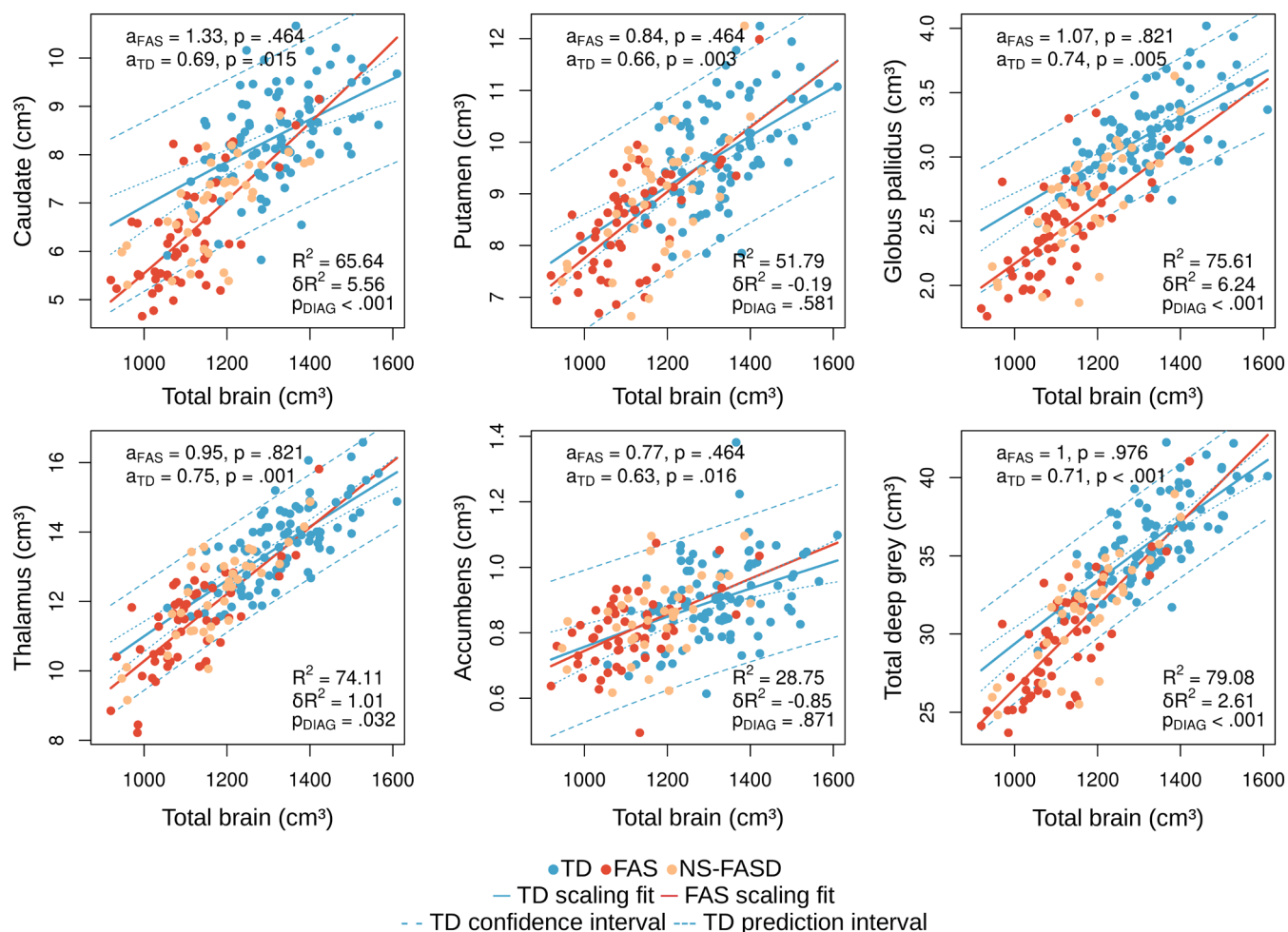


FIGURE 2 | Brain scaling charts of each deep grey matter nucleus volume and total deep grey matter volume with total brain volume in the fetal alcohol syndrome (FAS) and typically developing (TD) groups. NS-FASD, non-syndromic fetal alcohol spectrum disorder; a_{FAS}/a_{TD} , scaling coefficients and associated false discovery rate adjusted (FDR) p values; δR^2 , difference in adjusted coefficient of determination between models without and with FAS diagnosis as covariate and associated FDR-adjusted p value of model comparison.

notably increased the identification rate of individuals with a clinically promising FAS-like neuroanatomical phenotype, even without microcephaly.

4.1 | Deep Grey Matter Undersizing in FASD

4.1.1 | A Severe and Heterogeneous Pattern

The identification and characterization of a heterogeneous pattern of deep grey matter undersizing in FASD is difficult to confront with the literature. To date, only four studies have analyzed at least four major nuclei, excluding the accumbens, while accounting for total brain volume as a linear covariate (not using simple volumetric ratios) (Boateng et al. 2023; Gautam et al. 2015; Nardelli et al. 2011; Roussotte et al. 2012). Unlike our study, none of these tested the differences in the magnitude of volume reduction between nuclei. All four studies consistently found an over-reduction in globus pallidus volume, while the other nuclei showed significant over-reductions in only three studies and inconsistently so. Besides, only two studies have specifically examined the accumbens in individuals with FAS. Biffen et al. (2018) found no difference from controls after

adjusting for total intracranial volume (TIV), while Archibald et al. (2001) similarly reported no significant differences, controlling only for total subcortical volume. Interestingly, only one study used a pseudo-normative analysis similar to ours, examining residuals after modeling brain size effects. This study, which did not include the thalamus, found significant reductions in caudate and putamen volumes, with a near-significant trend for globus pallidus (Inkelis et al. 2020). While these results are not entirely in contradiction with ours, employing a more refined approach was necessary for drawing robust conclusions. As previously introduced, we aimed to address several methodological shortcomings, including patient group composition, modeling global brain size effects, and ensuring robustness across segmentation tools—factors that may contribute to inconsistencies among studies.

We first stratified the FASD sample into two subgroups to perform analyses separately and reduce the risk of diluting pathological effects by mixing certain (FAS) with only probable (NS-FASD) forms of the disease. While this does not contradict the continuum of volumetric impairments previously reported in FASD (Astley, Aylward, et al. 2009; Inkelis et al. 2020), it enhances the specificity and statistical power of our findings,

TABLE 2 | Mean percentages of volume deviation from typical scaling ($\sqrt[3]{\text{DTS}}$) of deep grey matter in the fetal alcohol syndrome (FAS) group, and the nonsyndromic fetal alcohol spectrum disorders (NS-FASD) group.

| ROIs | FAS | | NS-FASD | |
|------------------------|----------------------------|-----------------|----------------------------|-----------------|
| | $\sqrt[3]{\text{DTS}}$ (%) | CI 95% | $\sqrt[3]{\text{DTS}}$ (%) | CI 95% |
| Caudate | -12.61*** | [-15.47, -9.98] | -9.04*** | [-11.80, -6.23] |
| Putamen | -2.21* | [-4.34, -0.11] | -0.84 ^{ns} | [-4.16, 2.54] |
| Globus pallidus | -11.48*** | [-13.91, -9.23] | -7.13*** | [-10.02, -4.70] |
| Thalamus | -3.98** | [-5.76, -2.19] | -1.52 ^{ns} | [-3.47, 0.38] |
| Accumbens | -0.13 ^{ns} | [-2.79, 2.60] | 1.50 ^{ns} | [-2.20, 5.31] |
| Total deep grey matter | -6.10*** | [-7.86, -4.39] | -3.52*** | [-5.34, -1.73] |

Note: 95% CI, 95% confidence interval estimated by bootstrap resampling; false discovery rate adjusted p values: ns $p > 0.05$.

* $p \leq 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

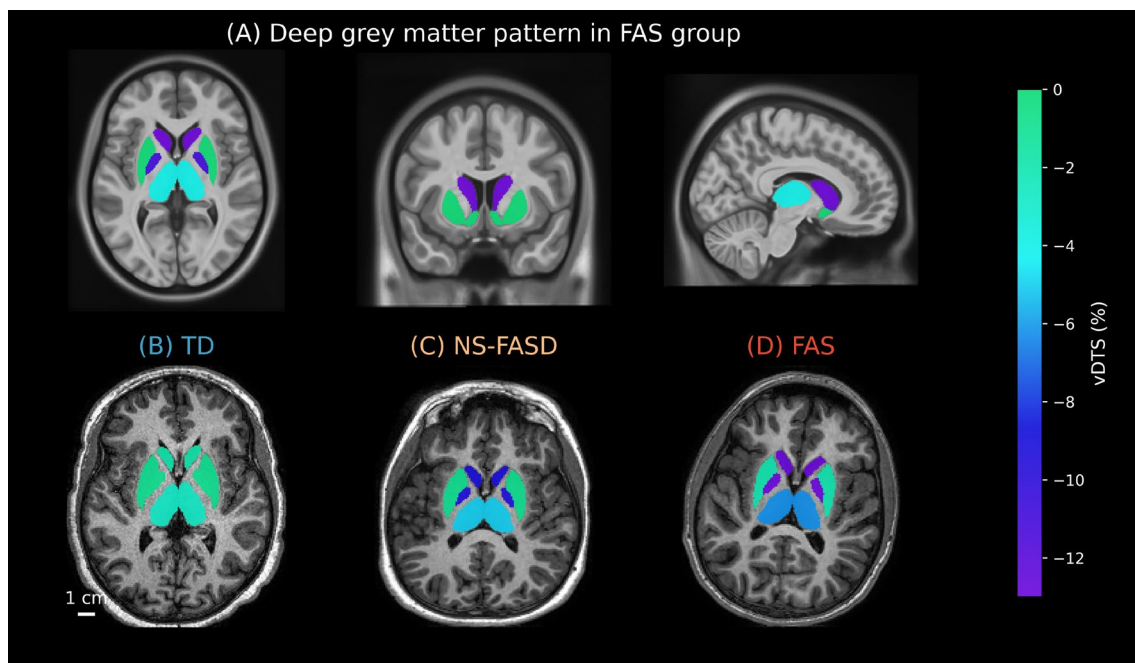


FIGURE 3 | Deep grey matter pattern of volume deviation from typical scaling (in %). (A) Representation in MNI standard space of the deep grey matter undersizing pattern in the fetal alcohol syndrome (FAS) group in sagittal, coronal, and axial planes, with data corresponding to those presented in Table 2. (B) Nuclear volume deviation from typical scaling ($\sqrt[3]{\text{DTS}}$) of the typically developing (TD) subject whose measurements are closest to the control group mean vector of $\sqrt[3]{\text{DTS}}$. (C) Nuclear $\sqrt[3]{\text{DTS}}$ of a subject with nonsyndromic fetal alcohol spectrum disorders (NS-FASD) whose measurements are closest to the FAS group mean vector of $\sqrt[3]{\text{DTS}}$. (D) Nuclear $\sqrt[3]{\text{DTS}}$ of the subject with FAS whose measurements are closest to the FAS group mean vector of $\sqrt[3]{\text{DTS}}$.

especially given the medium-sized sample (53 patients with FAS and 95 controls). Second, we used a power law to model the typical scaling relationship between deep grey matter volume and total brain volume. This non-linear approach has been shown to better fit data compared to proportional linear models (confirmed in our data, see Table S9), which are less biologically relevant and may bias estimates of group differences in contexts of abnormal growth (Germanaud et al. 2012; Toro et al. 2009). Specifically, considering the marked negative allometry of deep grey matter in controls (de Jong et al. 2017; Reardon et al. 2016), a proportional model based on relative volumes would have underestimated these differences, compromising sensitivity.

Third, we conducted a robustness analysis using two additional automated segmentation tools, previous studies having highlighted discrepancies between these tools and their versions (Biffen et al. 2020; Haddad et al. 2023; Sankar et al. 2017). In our data, the three tools showed strong correlations and high spatial overlap for all nuclei except the accumbens. The segmentation tools supported the finding of a severe undersizing of deep grey matter; however, they differed in estimating the magnitude of undersizing at both the global and nuclei levels. Notably, the order of severity across nuclei followed a similar pattern. The larger discrepancies of the accumbens aligned with the reference studies of volBrain and FSL-FIRST, which found less

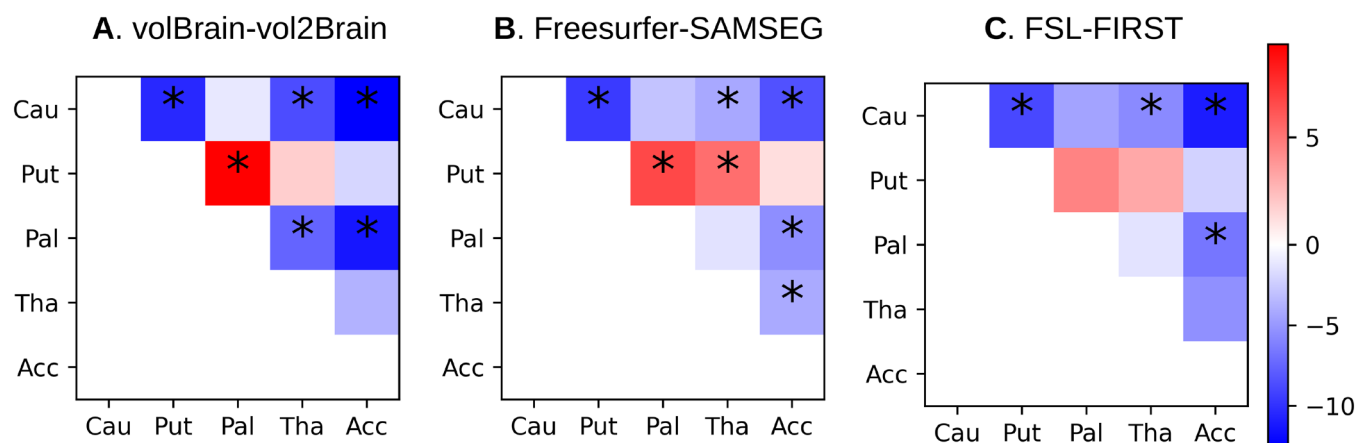


FIGURE 4 | Pairwise bootstrapped mean comparison matrices between deep grey matter volume deviations from typical scaling ($\sqrt{\text{DTS}}$) for each segmentation method. The color bar represents the magnitude and the direction of the $\sqrt{\text{DTS}}$ mean difference, and * indicates significant false discovery rate adjusted p values ($p \leq 0.05$). The bluer the cell, the smaller the nucleus $\sqrt{\text{DTS}}$ represented on the y-axis is compared with the one on the x-axis. Conversely, the redder the cell, the larger the nucleus $\sqrt{\text{DTS}}$ on the y-axis is compared to the one on the x-axis. Cau, Caudate; Put, Putamen; Pal, Globus pallidus; Tha, Thalamus; Acc, Accumbens.

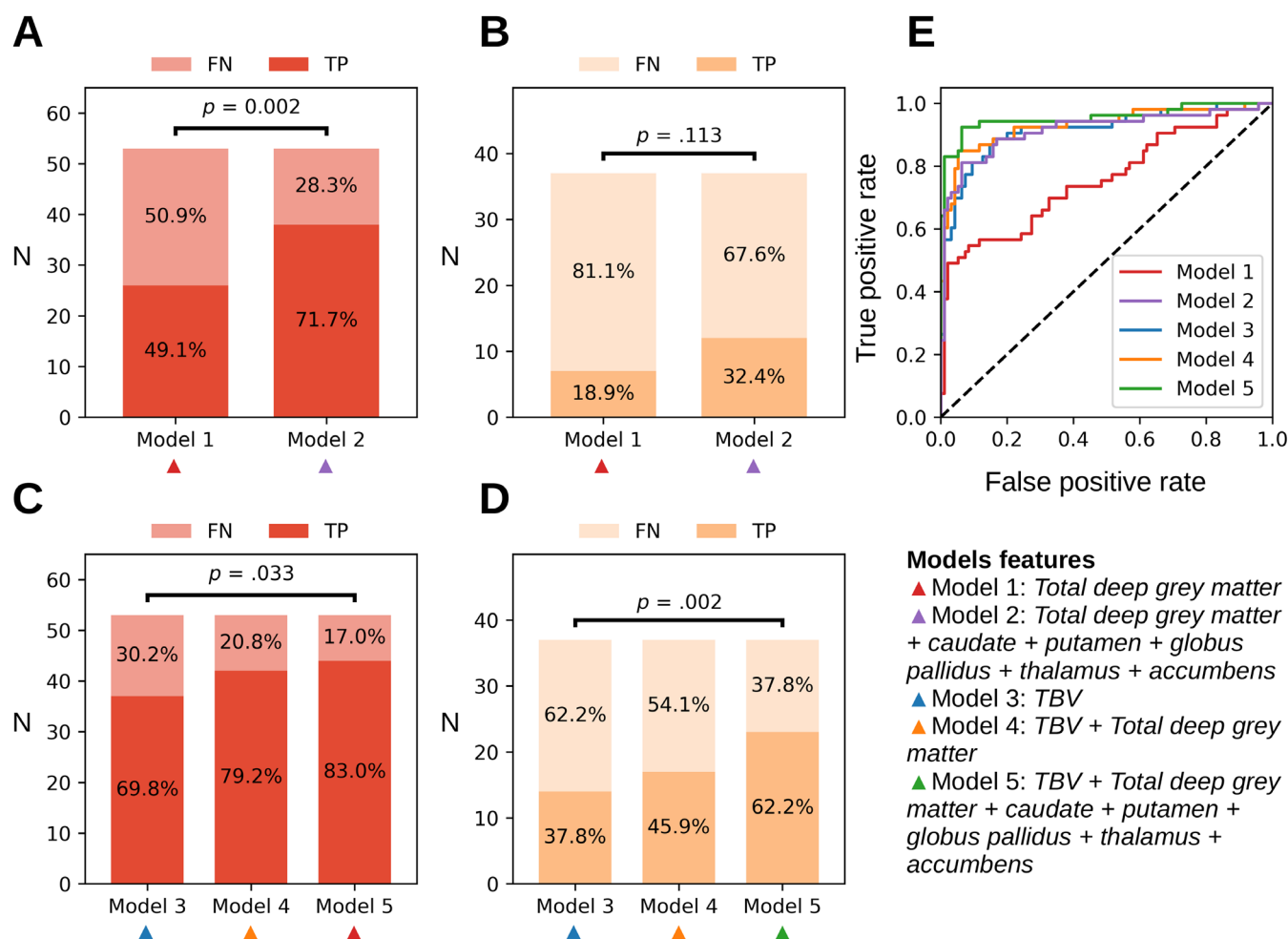


FIGURE 5 | Diagnostic classifiers applied to fetal alcohol syndrome (FAS) versus controls and FAS versus nonsyndromic fetal alcohol spectrum disorder (NS-FASD) groups. (A) Performance comparisons of 95% specificity classifiers based on total deep grey matter $\sqrt{\text{DTS}}$ with and without the individual nuclear $\sqrt{\text{DTS}}$ in the FAS sample and (B) NS-FASD sample. (C) Performance comparisons of 95% specificity classifiers based on total brain volume distance with and without the total deep grey matter and individual nuclear $\sqrt{\text{DTS}}$ in the FAS sample and (D) NS-FASD sample. (E) Receiver operation characteristic mean cross-validated curves of the five models. FN, false negative; TP, true positive; TBV, total brain volume.

robust segmentation of this structure compared to other deep grey matter structures (Manjón and Coupé 2016; Patenaude et al. 2011). This may be attributed to the neuroanatomical characteristics of the accumbens: it is a small structure with poorly defined boundaries, both in conventional T1-weighted imaging and anatomical dissection (Lucas-Neto et al. 2015; Neto et al. 2008). Nonetheless, given the potential role of the nucleus accumbens in the attentional and emotional phenotypes of FASD (Carmona et al. 2009; Zhuang et al. 2021), we included it in our study. Ultimately, confidence in the segmentation tools rests on scant published comparative data, which to date suggest a superior performance of volBrain, our chosen reference tool for clinical application (Hannoun et al. 2019; Manjón and Coupé 2016; Næss-Schmidt et al. 2016).

Ultimately, by employing a sample of patients with confirmed diagnoses, a biologically appropriate scaling model, and multiple segmentation tools, we provided strong evidence for a heterogeneous pattern of severe deep grey matter undersizing in FAS. The extent of volume reduction ranged from 0% to 12% below typical average volumes, with a 6% reduction for total deep grey matter. Notably, the caudate nucleus and globus pallidus showed the most consistent reductions across tools. However, variability between segmentation tools remains a major concern for clinical translation.

4.1.2 | Pathophysiological Relevance of This Pattern

The severe impairment of deep grey matter structures in FASD is consistent with the literature on animal models of prenatal alcohol exposure (PAE). Seminal work by Sulik et al. (1984) on mice first showed that prenatal alcohol exposure results in brain malformations, including hypoplasia of the basal ganglia. Subsequent studies, notably using computational MRI, replicated and extended these findings, suggesting increased vulnerability in specific deep grey matter structures such as the striatum, globus pallidus, and thalamus (Abbott et al. 2016; Coleman et al. 2012; Godin et al. 2010; Mattson et al. 1994). Furthermore, deep grey matter is not the first brain region where spatial patterns or gradients of undersizing were reported in FASD, as similar findings have been observed in the corpus callosum and cerebellum, across various size parameters (Fraize, Convert, et al. 2023; Fraize, Fischer, et al. 2023; Inkelis et al. 2020; Sullivan et al. 2020). However, the discrepancy between nuclei that share the same developmental origin (i.e., caudate and putamen)—which is even greater than that between those from different embryonic vesicles (i.e., caudate and thalamus)—raises questions about the underlying mechanisms of differential vulnerability to PAE within the striatum.

Several factors have been proposed and tested using animal models, including the timing of exposure (Coleman et al. 2012; Godin et al. 2010), dosage (Bonthius and West 1988; Ikonomidou et al. 2000) and mode of administration (Bonthius et al. 1988). To our knowledge, no animal study has directly compared the five nuclei investigated in our study, and those that examined the striatum did not subdivide it into its constituent nuclei, leaving gaps in our understanding of differential structural impacts. It is difficult to test the contribution of these factors to the spatial

heterogeneity observed in our clinical series since precise PAE data were unavailable. In addition, these factors may better explain inter-individual rather than group variability, except if the exposure conditions were highly uniform across subjects. Another source of variability could arise from individual genetic susceptibility (Streissguth and Dehaene 1993), which may not only contribute to differences between individuals but also influence the average pattern observed in sufficiently homogeneous clinical populations. Moreover, the effects of PAE on brain development could manifest later in childhood or adolescence (Wang et al. 2020). Our study, which covered ages 6 to 20, highlights an undersizing pattern that might differ in newborns or younger children, a population for which there is a notable lack of data (Lebel and Ware 2023). This potential delay may reflect effects that take time to develop (disruptions in the brain's developmental trajectory) or even feedback effects of adverse functioning on neuroanatomy. Finally, the influence of adverse environmental factors in the postnatal period, which are commonly associated in FASD patients (Lebel et al. 2019), should not be overlooked when interpreting our results, as they could further modulate brain development, compounding the effects of PAE and contributing to neuroanatomical variability (Andre et al. 2020).

In the end, PAE probably interacts with many factors in a complex process of reciprocal influences, affecting the overall development of the organism with short- and long-term consequences (Schneider et al. 2011). Despite significant progress and clear evidence of deep grey matter over-sensitivity to prenatal alcohol exposure from animal studies, the developmental toxicology of alcohol remains only partially understood, limiting our ability to draw out the pathophysiological relevance of the spatial heterogeneity of the deep grey matter impairment.

4.1.3 | Neuropsychological Relevance

The neurodevelopmental profiles of individuals with FASD are characterized by a high degree of inter-individual variability in cognitive and behavioral impairments, often spanning multiple domains (Astley, Olson, et al. 2009; Kodituwakku 2007). The most pronounced deficits frequently involve attention and executive functions (Chasnoff et al. 2015; Weyrauch et al. 2017), with a global impact on intelligence, although elementary reasoning skills are often preserved (Kerdreux et al. 2024). With regard to motor skills, although some studies suggest weaknesses in fine and gross motor skills (Doney et al. 2016; Lucas et al. 2014), epidemiological studies do not show a high prevalence of strict motor disorders associated with deep grey matter dysfunction (Chasnoff et al. 2015; Popova et al. 2016; Weyrauch et al. 2017).

One of the first nonstrictly motor anatomical-functional models of deep grey matter structures describes them as parallel circuits with the cortex, organized according to the functional cortical areas that project to the input nuclei. These projections are then topographically preserved along the circuit and project back to the same cortical areas. There is at least one cognitive loop connecting the dorsolateral prefrontal cortex to the caudate nucleus, an associative loop connecting the orbitofrontal cortex to the ventral striatum, and a motor loop connecting the (pre-)motor cortex to the putamen (Alexander et al. 1986; Tremblay et al. 2015). Consistent

with this model, functional MRI studies in humans have linked deep grey matter structures to various cognitive functions, such as the ventral striatum involvement in reinforcement learning (Calabro et al. 2023), and the roles of the dorsal striatum and globus pallidus in cognitive control and working memory (Guo et al. 2018; Zhuang et al. 2021; Rottschy et al. 2012). Furthermore, deep grey matter impairments are also observed in neurodevelopmental disorders without movement impairment at the group level. For instance, meta-analyses have found reduced volumes of the putamen, globus pallidus, and caudate in individuals with ADHD (Frodl and Skokauskas 2012; Norman et al. 2016). Considering this model, our results may suggest that the more pronounced impairment of the caudate nucleus and globus pallidus could account for the frequent attentional and executive difficulties profile frequently observed in the FASD population and in our series, while the less affected putamen might explain the absence of movement disorders. However, the relative preservation of the ventral striatum remains puzzling, given the frequent affect regulation difficulties (Temple et al. 2019).

Several studies have examined the relationship between deep grey matter impairment and overall functional severity in FASD (Fryer et al. 2012; Gimbel et al. 2024, 2025; Roussotte et al. 2012), but fine correlations are complicated by the fact that cortico-subcortical functional circuits are broad and complex. Indeed, studies reveal topographically organized di-synaptic connections between the striatum and cerebellum, involving both motor and non-motor areas (Hoshi et al. 2005). The observations suggest that the cortex, deep grey matter, and cerebellum form a densely interconnected network supporting both motor and non-motor functions (Bostan and Strick 2018). In addition, structural and functional connectivity studies have revealed a rostro-caudal gradient of associative to somatomotor connectivity with the cortex in each nucleus (Draganski et al. 2008; Greene et al. 2014; Jarbo and Verstynen 2015). The deep grey matter is part of a network distributed throughout the brain and cerebellum, with distinct motor, cognitive, and affective functional territories in each nucleus, making it hazardous to interpret the relationship between lesional (undersizing) and functional (cognitive) profiles on a rather coarse nucleus segmentation of the deep grey matter. To better understand the functional implications of a heterogeneous undersizing of the deep grey matter, it would be relevant to study it across the functional territories of each nucleus, using a parcellation based on structural or functional connectivity with the cortex (Draganski et al. 2008; Jarbo and Verstynen 2015) to try to identify more pronounced undersizing in specific functional territories.

4.2 | Deep Grey Matter-Based Diagnostic Classifier in FASD

4.2.1 | An Informative Undersizing Pattern That Boosts Classifiers

Adding the five nuclear $\sqrt{\text{DTS}}$ as predictors improved the classifier based on the global deep grey matter $\sqrt{\text{DTS}}$, showing that the nuclei provide additional diagnostic information. Notably, we already showed that a detailed regional allometric scaling profile in the cerebellum contains more discriminating information than the global sizing of the region in distinguishing patients with

FAS from typically developing individuals with high specificity (set > 95%) (Fraize, Fischer, et al. 2023). Here, we go even further by demonstrating a significant add-on effect on a reference classifier based on the sole neuroanatomical criterion endorsed in diagnostic guidelines—that is, deviation from the reference mean cerebral volume (microcephaly) (Astley and Clarren 2000; Cook et al. 2016; Hoyme et al. 2016). Indeed, the nuclear predictors remained informative even when combined with the already highly discriminating presence or absence of microcephaly, present in approximately 70% of our patients with FAS. The benefit of the nuclear predictors became even more apparent in our NS-FASD sample: applied to this population where microcephaly is less common (30%), the multiple-predictors classifier trained on the FAS and typically developing samples identified more than 1.5 times as many NS-FASD patients with a FAS-like phenotype as the reference classifier, meaning that almost one third of the NS-FASD group did not have microcephaly but still had a deep grey undersizing profile compatible with the disease.

4.2.2 | Deep Grey Signature Transfer and Classifier Design

Such a result on NS-FASD is consistent with previous studies (Astley, Aylward, et al. 2009; Boateng et al. 2023) showing that these patients have an intermediate neuroanatomical phenotype compared to patients with FAS at the group level, but may present similarly intense features at the individual level. More specifically, we demonstrated that a distinctive neuroanatomical signature of FAS can be detected and learned by a classifier, enabling the identification of NS-FASD patients who, by definition, lack the specific facial features but exhibit neuroanatomical markers of FAS. This approach shares similarities with the transfer learning method (Dufumier et al. 2024), where a classifier is trained on patients with a definitive diagnostic form of the pathology and then applied to an unseen, out-of-domain sample of patients with a nonspecific form of the condition, characterized by intermediate cerebral impairments. Unlike conventional transfer learning, this process omits the fine-tuning stage, where the model would typically be re-trained on the new dataset. This out-of-domain classifier application has already been employed in biomarker searches for psychiatric and neurodegenerative disorders. Notable examples include the identification of a volumetric brain signature of schizophrenia in healthy relatives (Fan et al. 2008) and the prediction of mild cognitive impairment progression to Alzheimer's disease (Cui et al. 2011). To our knowledge, no study in the field of FASD neuroimaging has used such an out-of-domain application. Recently, Das et al. (2024) used a conventional deep transfer learning method, but their aim was rather to compensate for the relatively small size of their PAE and control samples by transferring the knowledge of low-level features, learned by an age-and-sex predicting model pre-trained on the UK-Biobank cohort (Gong et al. 2021), to their diagnostic classifier.

The best approach to develop diagnostic classifiers based on MRI features remains an open question. Although unsupervised deep learning strategies leveraging big datasets have shown promising results (Rakić et al. 2020), these methods are not applicable in the specific context of FASD, where access to large volumes of data is limited. This is why we chose a tailored, potentially

more parsimonious strategy. We used a moderate size sample, including a patient group with a high prevalence of severe forms of the disease. These patients underwent a rigorous diagnostic procedure, including a systematic search for differential diagnoses based on genetic and biological assessments, to reduce etiological noise and maximize contrast with the control group. We then selected a linear statistical model (logistic regression), suitable for binary diagnostic classification, which has been shown to be effective notably in phenotype prediction studies based on neuroanatomical features, including in contexts of small samples (Dufumier et al. 2024; Schulz et al. 2020, 2024). Finally, for feature engineering, we followed the normative approach, widely advocated by Marquand et al. (2016) for its relevance in diagnostic classification. This approach has recently been used to enhance prenatal alcohol exposure diagnostic sensitivity based on growth charts of white matter, cortical, and subcortical grey matter (Gimbel et al. 2024). However, unlike these studies, we advocate for using total brain volume rather than age as the reference (i.e., normative scaling analysis) in specific cases of brain growth restriction to maximize the discrimination power between patients and controls.

4.2.3 | Clinical Relevance

Thirty years have passed since the seminal work of Mattson et al. (1994, 1996) suggesting an over-reduction of deep grey matter in FASD, in a context where other quantitative focal anomalies were beginning to be described. But since then, no neuroanatomical criterion other than microcephaly has been formally incorporated into diagnostic guidelines (Astley and Clarren 2000; Cook et al. 2016; Hoyme et al. 2016). However, the clinical criteria for positive diagnosis of non-syndromic forms of FASD remain clearly inadequate, arguing for additional sources of evidence in the absence of FAS. Our study investigated and demonstrated the added diagnostic value of normative deep grey matter features in comparison to microcephaly and showed how these features can be used concretely to identify patients with NS-FASD who have a phenotype compatible with FAS, thus for whom the causal link to PAE is arguably all the more likely. In a previous work, we proposed a similar use of a diagnostic classifier based on the gradient of cerebellar undersizing, but without testing the contribution of this gradient to the diagnosis when combined with total brain volume reduction (Fraize, Fischer, et al. 2023). Surprisingly enough, there have been very few attempts to bridge the diagnosis gap in FASD with MRI-based classifiers. Little and Beaulieu (2020) employed a classifier based on regional brain volumes to identify the regions that contributed the most to FASD classification. But by combining multiple subtypes of the disease down to undiagnosed PAE subjects, their goal was not to explore the quasi-transfer of learning from the syndrome to non-specific or even nondiagnosed forms. Additionally, they did not include total brain volume as a predictor, which could account for the weighting of certain features in the classification.

The absence of at least a third-party comparison group to emulate differential diagnosis situations, for instance individuals with microcephaly and/or neurodevelopmental disorders not caused by PAE, temporarily limits our study to proof-of-concept. Including such a group at least as a test sample, if

not in the training population, would be a key step to clinical translation, to ensure that the FAS deep grey matter signature is indeed specific to the disease and not only efficient in distinguishing any atypical neuroanatomy. In other words, for the time being, the classifier mainly identified atypical anatomies very effectively based on what it has learned from FAS, with no guarantee that this signature distinguishes FASD from other atypical situations. Nevertheless, as part of our iterative research approach, each set of regional features may enrich the neuroanatomical signature and increase our chances of identifying a whole-brain dimensional profile that is truly specific to FASD—provided that the risk of overfitting is adequately controlled in the end.

Eventually, our results support the idea that computational imaging and normative scaling analysis can provide MRI-based diagnostic classifiers capable of advancing FASD diagnostic tools to improve reliability in non-syndromic forms. This direction was already proposed when we first introduced a structured neuroanatomical score, based on common FAS abnormalities revealed by clinical imaging (brain size, thickness of the corpus callosum, height and foliation of the cerebellar vermis), to be added to the classical diagnostic tree (Fraize et al. 2022).

4.3 | Built-In Limitations

A primary limitation of our study relates to its retrospective design and the fact that most of the participants with FASD were adopted children, making it difficult to accurately report the precise amount of PAE, although we applied appropriate diagnostic thresholds whenever possible (Cook et al. 2016; Hoyme et al. 2016). Additional information on maternal polysubstance use—common in mothers with alcohol use disorder (Tran et al. 2023)—and other adverse postnatal environmental factors that may interact with the effects of PAE on neurodevelopment (Andre et al. 2020; Lebel et al. 2019) was also challenging to collect.

Age may also be a confounding factor in our analyses since our sample included patients aged 6 to 20 years a broad developmental period marked by complex brain growth and maturation. Unlike cortical volumes, which typically show a downward trajectory during this age range, deep grey matter structures exhibit slower and more gradual growth (Raznahan et al. 2014). This differential developmental trajectory likely explains the limited effects of age we observed. Given the need for larger datasets to fully capture the small age-related changes in this period, we opted for the linear residualization of age effects when significant, close to considering age as a covariate without interaction in classical regression models (Nardelli et al. 2011; Roussotte et al. 2012).

An additional limitation of our analysis design is the lack of sex-specific subgroup analyses. Sexual dimorphism in deep grey matter is well documented, with males generally exhibiting larger volumes and females reaching peak growth earlier (Raznahan et al. 2014). In FASD, volumetric differences between patients and typically developing controls tend to be more pronounced in males, particularly in regions such as the striatum, globus pallidus, and thalamus (Inkelis et al. 2020; Treit et al. 2017).

Sex-specific analyses could have provided valuable insights, but as we already argued in other normative studies within the same dataset, we prioritized statistical power provided by larger numbers by adjusting for sex in the volumetric analyses. While many studies controlled for sex using linear methods (Biffen et al. 2018; Boateng et al. 2023; Nardelli et al. 2011), we applied a batch effect correction method (Fortin et al. 2018), consistent with prior work (Fraize, Fischer, et al. 2023). This approach is statistically sound and provides a more efficient and robust prior account of sex-related effects than simple linear residualization, by accounting not only for additive but also for multiplicative effects (such as variance differences between sexes), and by using robust parameter estimation based on an empirical Bayesian framework. We consider the effectiveness of this adjustment trustworthy (very unlikely to have corrupted the results), as post hoc analyses implementing linear residualization of sex (Table S10), or robustness analyses that did not control for age and sex (Table S2), yielded very similar results. However, adjusting for sex in this manner may obscure complex interactions between PAE and sex, potentially limiting our ability to detect more subtle, sex-specific diagnostic markers in FASD. Indeed, the development of sex-specific diagnostic classifiers appears increasingly important, especially considering research by Little and Beaulieu (2020), which demonstrated that the most contributive brain regions for classification differ by sex: deep grey matter structures were most contributive in boys, while cortical regions played a predominant role in girls.

5 | Conclusion

Our study revealed a heterogeneous pattern of deep grey matter allometric undersizing in FASD, providing diagnostic features beyond global deep grey matter and even total brain volume reductions to feed classifiers. The deep grey matter-based neuroanatomical FAS signature learned by these classifiers was recognized in nearly two-thirds of non-syndromic FASD, where it can be considered an additional piece of evidence strengthening the likelihood of the causal link with alcohol exposure, even in the absence of microcephaly. By providing new evidence for the diagnostic relevance of neuroanatomical MRI-based features, this work represents a meaningful step toward integrating such biomarkers into FASD clinical guidelines thanks to computational classifiers.

Acknowledgments

This study was supported by the Agence Nationale de la Recherche (ANR-19-CE17-0028-01), a grant from the French government managed by the Agence Nationale de la Recherche as part of the France 2030 program (ANR-23-IAHU-0010), and the Institut pour la Recherche en Santé Publique (IRES-19-ADDICTIONS-08). We thank the volunteers, patients, and families, the French supportive association for FASD-affected families “Vivre avec le SAF”, and Elizabeth Rowley-Jolivet for English proofreading.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

- Abbott, C. W., O. O. Kozanian, J. Kanaan, K. M. Wendel, and K. J. Huffman. 2016. “The Impact of Prenatal Ethanol Exposure on Neuroanatomical and Behavioral Development in Mice.” *Alcoholism, Clinical and Experimental Research* 40, no. 1: 122–133. <https://doi.org/10.1111/acer.12936>.
- Akudjedu, T. N., L. Nabulsi, M. Makelyte, et al. 2018. “A Comparative Study of Segmentation Techniques for the Quantification of Brain Subcortical Volume.” *Brain Imaging and Behavior* 12, no. 6: 1678–1695. <https://doi.org/10.1007/s11682-018-9835-y>.
- Alexander, G. E., M. R. DeLong, and P. L. Strick. 1986. “Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex.” *Annual Review of Neuroscience* 9, no. 1: 357–381.
- Andre, Q. R., C. A. McMorris, P. Kar, et al. 2020. “Different Brain Profiles in Children With Prenatal Alcohol Exposure With or Without Early Adverse Exposures.” *Human Brain Mapping* 41, no. 15: 4375–4385. <https://doi.org/10.1002/hbm.25130>.
- Archibald, S. L., C. Fennema-Notestine, A. Gamst, E. P. Riley, S. N. Mattson, and T. L. Jernigan. 2001. “Brain Dymorphology in Individuals With Severe Prenatal Alcohol Exposure.” *Developmental Medicine and Child Neurology* 43, no. 3: 148. <https://doi.org/10.1017/S0012162201000299>.
- Astley, S. J., and S. K. Clarren. 2000. “Diagnosing the Full Spectrum of Fetal Alcohol-Exposed Individuals: Introducing the 4-Digit Diagnostic Code.” *Alcohol and Alcoholism* 35, no. 4: 400–410. <https://doi.org/10.1093/alcalc/35.4.400>.
- Astley, S. J., E. H. Aylward, H. C. Olson, et al. 2009. “Magnetic Resonance Imaging Outcomes From a Comprehensive Magnetic Resonance Study of Children With Fetal Alcohol Spectrum Disorders.” *Alcoholism: Clinical and Experimental Research* 33, no. 10: 1671–1689. <https://doi.org/10.1111/j.1530-0277.2009.01004.x>.
- Astley, S. J., H. C. Olson, K. Kerns, et al. 2009. “Neuropsychological and Behavioral Outcomes From a Comprehensive Magnetic Resonance Study of Children With Fetal Alcohol Spectrum Disorders.” *Canadian Journal of Clinical Pharmacology=Journal Canadien de Pharmacologie Clinique* 16, no. 1: e178–e201.
- Benjamini, Y., and Y. Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 57, no. 1: 289–300.
- Biffen, S. C., C. M. R. Warton, N. C. Dodge, et al. 2020. “Validity of Automated FreeSurfer Segmentation Compared to Manual Tracing in Detecting Prenatal Alcohol Exposure-Related Subcortical and Corpus Callosal Alterations in 9- to 11-Year-Old Children.” *NeuroImage Clinical* 28: 102368. <https://doi.org/10.1016/j.nicl.2020.102368>.
- Biffen, S. C., C. M. R. Warton, N. M. Lindinger, et al. 2018. “Reductions in Corpus Callosum Volume Partially Mediate Effects of Prenatal Alcohol Exposure on IQ.” *Frontiers in Neuroanatomy* 11: 132. <https://doi.org/10.3389/fnana.2017.00132>.
- Boateng, T., K. Beauchamp, F. Torres, et al. 2023. “Brain Structural Differences in Children With Fetal Alcohol Spectrum Disorder and Its Subtypes.” *Frontiers in Neuroscience* 17: 1152038. <https://doi.org/10.3389/fnins.2023.1152038>.
- Bonthius, D. J., and J. R. West. 1988. “Blood Alcohol Concentration and Microencephaly: A Dose-Response Study in the Neonatal Rat.” *Teratology* 37, no. 3: 223–231. <https://doi.org/10.1002/tera.1420370307>.
- Bonthius, D. J., C. R. Goodlett, and J. R. West. 1988. “Blood Alcohol Concentration and Severity of Microencephaly in Neonatal Rats Depend on the Pattern of Alcohol Administration.” *Alcohol* 5, no. 3: 209–214. [https://doi.org/10.1016/0741-8329\(88\)90054-7](https://doi.org/10.1016/0741-8329(88)90054-7).
- Bostan, A. C., and P. L. Strick. 2018. “The Basal Ganglia and the Cerebellum: Nodes in an Integrated Network.” *Nature Reviews*

- Neuroscience* 19, no. 6: 338–350. <https://doi.org/10.1038/s41583-018-0002-7>.
- Bouyeure, A., D. Germanaud, D. Bekha, et al. 2018. “Three-Dimensional Probabilistic Maps of Mesial Temporal Lobe Structures in Children and Adolescents’ Brains.” *Frontiers in Neuroanatomy* 12: 98. <https://doi.org/10.3389/fnana.2018.00098>.
- Calabro, F. J., D. F. Montez, B. Larsen, et al. 2023. “Striatal Dopamine Supports Reward Expectation and Learning: A Simultaneous PET/fMRI Study.” *NeuroImage* 267: 119831. <https://doi.org/10.1016/j.neuroimage.2022.119831>.
- Carmona, S., E. Proal, E. A. Hoekzema, et al. 2009. “Ventral Striatal Reductions Underpin Symptoms of Hyperactivity and Impulsivity in Attention-Deficit/Hyperactivity Disorder.” *Biological Psychiatry* 66, no. 10: 972–977. <https://doi.org/10.1016/j.biopsych.2009.05.013>.
- Chasnoff, I. J., A. M. Wells, and L. King. 2015. “Misdiagnosis and Missed Diagnoses in Foster and Adopted Children With Prenatal Alcohol Exposure.” *Pediatrics* 135, no. 2: 264–270. <https://doi.org/10.1542/peds.2014-2171>.
- Chen, X., C. D. Coles, M. E. Lynch, and X. Hu. 2012. “Understanding Specific Effects of Prenatal Alcohol Exposure on Brain Structure in Young Adults.” *Human Brain Mapping* 33, no. 7: 1663–1676. <https://doi.org/10.1002/hbm.21313>.
- Coleman, L. G., I. Oguz, J. Lee, M. Styner, and F. T. Crews. 2012. “Postnatal Day 7 Ethanol Treatment Causes Persistent Reductions in Adult Mouse Brain Volume and Cortical Neurons With Sex Specific Effects on Neurogenesis.” *Alcohol* 46, no. 6: 603–612. <https://doi.org/10.1016/j.alcohol.2012.01.003>.
- Cook, J. L., C. R. Green, C. M. Lilley, et al. 2016. “Fetal Alcohol Spectrum Disorder: A Guideline for Diagnosis Across the Lifespan.” *Canadian Medical Association Journal* 188, no. 3: 191–197. <https://doi.org/10.1503/cmaj.141593>.
- Cui, Y., B. Liu, S. Luo, et al. 2011. “Identification of Conversion From Mild Cognitive Impairment to Alzheimer’s Disease Using Multivariate Predictors.” *PLoS One* 6, no. 7: e21896. <https://doi.org/10.1371/journal.pone.0021896>.
- Das, A., K. Duarte, C. Lebel, and M. Bento. 2024. “Deep Learning for Detecting Prenatal Alcohol Exposure in Pediatric Brain MRI: A Transfer Learning Approach With Explainability Insights.” *Frontiers in Computational Neuroscience* 18: 1434421. <https://doi.org/10.3389/fncom.2024.1434421>.
- de Jong, L. W., J.-S. Vidal, L. E. Forsberg, et al. 2017. “Allometric Scaling of Brain Regions to Intra-Cranial Volume: An Epidemiological MRI Study.” *Human Brain Mapping* 38, no. 1: 151–164. <https://doi.org/10.1002/hbm.23351>.
- Donald, K. A., E. Eastman, F. M. Howells, et al. 2015. “Neuroimaging Effects of Prenatal Alcohol Exposure on the Developing Human Brain: A Magnetic Resonance Imaging Review.” *Acta Neuropsychiatrica* 27, no. 5: 251–269. <https://doi.org/10.1017/neu.2015.12>.
- Doney, R., B. R. Lucas, R. E. Watkins, et al. 2016. “Visual-Motor Integration, Visual Perception, and Fine Motor Coordination in a Population of Children With High Levels of Fetal Alcohol Spectrum Disorder.” *Research in Developmental Disabilities* 55: 346–357. <https://doi.org/10.1016/j.ridd.2016.05.009>.
- Draganski, B., F. Kherif, S. Klöppel, et al. 2008. “Evidence for Segregated and Integrative Connectivity Patterns in the Human Basal Ganglia.” *Journal of Neuroscience* 28, no. 28: 7143–7152. <https://doi.org/10.1523/JNEUROSCI.1486-08.2008>.
- Dufumier, B., P. Gori, S. Petiton, et al. 2024. “Exploring the Potential of Representation and Transfer Learning for Anatomical Neuroimaging: Application to Psychiatry.” *NeuroImage* 296: 120665. <https://doi.org/10.1016/j.neuroimage.2024.120665>.
- Fan, Y., R. E. Gur, R. C. Gur, et al. 2008. “Unaffected Family Members and Schizophrenia Patients Share Brain Structure Patterns: A High-Dimensional Pattern Classification Study.” *Biological Psychiatry* 63, no. 1: 118–124. <https://doi.org/10.1016/j.biopsych.2007.03.015>.
- Flak, A. L., S. Su, J. Bertrand, C. H. Denny, U. S. Kesmodel, and M. E. Cogswell. 2014. “The Association of Mild, Moderate, and Binge Prenatal Alcohol Exposure and Child Neuropsychological Outcomes: A Meta-Analysis.” *Alcoholism, Clinical and Experimental Research* 38, no. 1: 214–226. <https://doi.org/10.1111/acer.12214>.
- Fortin, J.-P., N. Cullen, Y. I. Sheline, et al. 2018. “Harmonization of Cortical Thickness Measurements Across Scanners and Sites.” *NeuroImage* 167: 104–120. <https://doi.org/10.1016/j.neuroimage.2017.11.024>.
- Fraize, J., C. Fischer, M. Elmaleh-Bergès, et al. 2023. “Enhancing Fetal Alcohol Spectrum Disorders Diagnosis With a Classifier Based on the Intracerebellar Gradient of Volumetric Undersizing.” *Human Brain Mapping* 44: 4321–4336. <https://doi.org/10.1002/hbm.26348>.
- Fraize, J., G. Convert, Y. Leprince, et al. 2023. “Mapping Corpus Callosum Surface Reduction in Fetal Alcohol Spectrum Disorders With Sulci and Connectivity-Based Parcellation.” *Frontiers in Neuroscience* 17: 1188367. <https://doi.org/10.3389/fnins.2023.1188367>.
- Fraize, J., P. Garzón, A. Ntorkou, et al. 2022. “Combining Neuroanatomical Features to Support Diagnosis of Fetal Alcohol Spectrum Disorders.” *Developmental Medicine and Child Neurology* 65, no. 4: 551–562. <https://doi.org/10.1111/dmcn.15411>.
- Fraize, J., Y. Leprince, M. Elmaleh-Bergès, et al. 2023. “Spectral-Based Thickness Profiling of the Corpus Callosum Enhances Anomaly Detection in Fetal Alcohol Spectrum Disorders.” *Frontiers in Neuroscience* 17: 1289013. <https://doi.org/10.3389/fnins.2023.1289013>.
- Frodil, T., and N. Skokauskas. 2012. “Meta-Analysis of Structural MRI Studies in Children and Adults With Attention Deficit Hyperactivity Disorder Indicates Treatment Effects.” *Acta Psychiatrica Scandinavica* 125, no. 2: 114–126. <https://doi.org/10.1111/j.1600-0447.2011.01786.x>.
- Fryer, S. L., S. N. Mattson, T. L. Jernigan, S. L. Archibald, K. L. Jones, and E. P. Riley. 2012. “Caudate Volume Predicts Neurocognitive Performance in Youth With Heavy Prenatal Alcohol Exposure.” *Alcoholism: Clinical and Experimental Research* 36, no. 11: 1932–1941. <https://doi.org/10.1111/j.1530-0277.2012.01811.x>.
- Gautam, P., C. Lebel, K. L. Narr, et al. 2015. “Volume Changes and Brain-Behavior Relationships in White Matter and Subcortical Gray Matter in Children With Prenatal Alcohol Exposure.” *Human Brain Mapping* 36, no. 6: 2318–2329. <https://doi.org/10.1002/hbm.22772>.
- Germanaud, D., J. Lefèvre, C. Fischer, et al. 2014. “Simplified Gyrification Pattern in Severe Developmental Microcephalies? New Insights From Allometric Modeling for Spatial and Spectral Analysis of Gyrification.” *NeuroImage* 102: 317–331. <https://doi.org/10.1016/j.neuroimage.2014.07.057>.
- Germanaud, D., J. Lefèvre, R. Toro, et al. 2012. “Larger Is Twistier: Spectral Analysis of Gyrification (SPANGY) Applied to Adult Brain Size Polymorphism.” *NeuroImage* 63, no. 3: 1257–1272. <https://doi.org/10.1016/j.neuroimage.2012.07.053>.
- Gimbel, B. A., D. J. Roediger, A. M. Ernst, et al. 2024. “Normative Magnetic Resonance Imaging Data Increase the Sensitivity to Brain Volume Abnormalities in the Classification of Fetal Alcohol Spectrum Disorder.” *Journal of Pediatrics* 266: 113868. <https://doi.org/10.1016/j.jpeds.2023.113868>.
- Gimbel, B. A., D. J. Roediger, M. E. Anthony, et al. 2025. “Normative Modeling of Brain MRI Data Identifies Small Subcortical Volumes and Associations With Cognitive Function in Youth With Fetal Alcohol Spectrum Disorder (FASD).” *NeuroImage: Clinical* 45: 103722. <https://doi.org/10.1016/j.nicl.2024.103722>.

- Godin, E. A., S. K. O'Leary-Moore, A. A. Khan, et al. 2010. "Magnetic Resonance Microscopy Defines Ethanol-Induced Brain Abnormalities in Prenatal Mice: Effects of Acute Insult on Gestational Day 7." *Alcoholism: Clinical and Experimental Research* 34, no. 1: 98–111. <https://doi.org/10.1111/j.1530-0277.2009.01071.x>.
- Gong, W., C. F. Beckmann, A. Vedaldi, S. M. Smith, and H. Peng. 2021. "Optimising a Simple Fully Convolutional Network for Accurate Brain Age Prediction in the PAC 2019 Challenge." *Frontiers in Psychiatry* 12: 627996. <https://doi.org/10.3389/fpsy.2021.627996>.
- Greene, D. J., T. O. Laumann, J. W. Dubis, et al. 2014. "Developmental Changes in the Organization of Functional Connections Between the Basal Ganglia and Cerebral Cortex." *Journal of Neuroscience* 34, no. 17: 5842–5854. <https://doi.org/10.1523/JNEUROSCI.3069-13.2014>.
- Guo, Y., T. W. Schmitz, M. Mur, C. S. Ferreira, and M. C. Anderson. 2018. "A Supramodal Role of the Basal Ganglia in Memory and Motor Inhibition: Meta-Analytic Evidence." *Neuropsychologia* 108: 117–134. <https://doi.org/10.1016/j.neuropsychologia.2017.11.033>.
- Haddad, E., F. Pizzagalli, A. H. Zhu, et al. 2023. "Multisite Test–Retest Reliability and Compatibility of Brain Metrics Derived From FreeSurfer Versions 7.1, 6.0, and 5.3." *Human Brain Mapping* 44, no. 4: 1515–1532. <https://doi.org/10.1002/hbm.26147>.
- Hannoun, S., R. Tutunji, M. El Homsy, S. Saaybi, and R. Hourani. 2019. "Automatic Thalamus Segmentation on Unenhanced 3D T1 Weighted Images: Comparison of Publicly Available Segmentation Methods in a Pediatric Population." *Neuroinformatics* 17, no. 3: 443–450. <https://doi.org/10.1007/s12021-018-9408-7>.
- Hoshi, E., L. Tremblay, J. Féger, P. L. Carras, and P. L. Strick. 2005. "The Cerebellum Communicates With the Basal Ganglia." *Nature Neuroscience* 8, no. 11: 1491–1493. <https://doi.org/10.1038/nn1544>.
- Hoyme, H. E., W. O. Kalberg, A. J. Elliott, et al. 2016. "Updated Clinical Guidelines for Diagnosing Fetal Alcohol Spectrum Disorders." *Pediatrics* 138, no. 2: 20.
- Ikonomidou, C., P. Bittigau, M. J. Ishimaru, et al. 2000. "Ethanol-Induced Apoptotic Neurodegeneration and Fetal Alcohol Syndrome." *Science* 287, no. 5455: 1056–1060. <https://doi.org/10.1126/science.287.5455.1056>.
- Inkelis, S. M., E. M. Moore, A. Bischoff-Grethe, and E. P. Riley. 2020. "Neurodevelopment in Adolescents and Adults With Fetal Alcohol Spectrum Disorders (FASD): A Magnetic Resonance Region of Interest Analysis." *Brain Research* 1732: 146654. <https://doi.org/10.1016/j.brainres.2020.146654>.
- Jarbo, K., and T. D. Verstynen. 2015. "Converging Structural and Functional Connectivity of Orbitofrontal, Dorsolateral Prefrontal, and Posterior Parietal Cortex in the Human Striatum." *Journal of Neuroscience* 35, no. 9: 3865–3878. <https://doi.org/10.1523/JNEUROSCI.2636-14.2015>.
- Jia, J. 2020. "A Package to Compute Segmentation Metrics: Seg-Metrics [Python]." https://github.com/Ordgod/segmentation_metrics.
- Kerdreux, E., J. Fraize, P. Garzón, et al. 2024. "Questioning Cognitive Heterogeneity and Intellectual Functioning in Fetal Alcohol Spectrum Disorders From the Wechsler Intelligence Scale for Children." *Clinical Neuropsychologist* 38, no. 5: 1109–1132. <https://doi.org/10.1080/13854046.2023.2281703>.
- Kodituwakku, P. W. 2007. "Defining the Behavioral Phenotype in Children With Fetal Alcohol Spectrum Disorders: A Review." *Neuroscience and Biobehavioral Reviews* 31, no. 2: 192–201. <https://doi.org/10.1016/j.neubiorev.2006.06.020>.
- Lebel, C., and A. Ware. 2023. "Magnetic Resonance Imaging in Fetal Alcohol Spectrum Disorder (FASD)." In *Neurodevelopmental Pediatrics: Genetic and Environmental Influences*, edited by D. D. Eisenstat, D. Goldowitz, T. F. Oberlander, and J. Y. Yager, 397–407. Springer International Publishing. https://doi.org/10.1007/978-3-031-20792-1_25.
- Lebel, C., C. A. McMorris, P. Kar, et al. 2019. "Characterizing Adverse Prenatal and Postnatal Experiences in Children." *Birth Defects Research* 111, no. 12: 848–858. <https://doi.org/10.1002/bdr2.1464>.
- Little, G., and C. Beaulieu. 2020. "Multivariate Models of Brain Volume for Identification of Children and Adolescents With Fetal Alcohol Spectrum Disorder." *Human Brain Mapping* 41, no. 5: 1181–1194. <https://doi.org/10.1002/hbm.24867>.
- Liu, D., H. J. Johnson, J. D. Long, V. A. Magnotta, and J. S. Paulsen. 2014. "The Power-Proportion Method for Intracranial Volume Correction in Volumetric Imaging Analysis." *Frontiers in Neuroscience* 8: 356. <https://doi.org/10.3389/fnins.2014.00356>.
- Lucas-Neto, L., S. Reimão, E. Oliveira, et al. 2015. "Advanced MR Imaging of the Human Nucleus Accumbens—Additional Guiding Tool for Deep Brain Stimulation." *Neuromodulation: Technology at the Neural Interface* 18, no. 5: 341–348. <https://doi.org/10.1111/ner.12289>.
- Lucas, B. R., J. Latimer, R. Z. Pinto, et al. 2014. "Gross Motor Deficits in Children Prenatally Exposed to Alcohol: A Meta-Analysis." *Pediatrics* 134, no. 1: e192–e209. <https://doi.org/10.1542/peds.2013-3733>.
- Manjón, J. V., and P. Coupé. 2016. "volBrain: An Online MRI Brain Volumetry System." *Frontiers in Neuroinformatics* 10: 30. <https://doi.org/10.3389/fninf.2016.00030>.
- Manjón, J. V., J. E. Romero, R. Vivo-Hernando, et al. 2022. "vol2Brain: A New Online Pipeline for Whole Brain MRI Analysis." *Frontiers in Neuroinformatics* 16: 862805. <https://doi.org/10.3389/fninf.2022.862805>.
- Marquand, A. F., I. Rezek, J. Buitelaar, and C. F. Beckmann. 2016. "Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies." *Biological Psychiatry* 80, no. 7: 552–561. <https://doi.org/10.1016/j.biopsych.2015.12.023>.
- Mattson, S. N., E. P. Riley, E. R. Sowell, T. L. Jernigan, D. F. Sobel, and K. L. Jones. 1996. "A Decrease in the Size of the Basal Ganglia in Children With Fetal Alcohol Syndrome." *Alcoholism: Clinical and Experimental Research* 20, no. 6: 1088–1093. <https://doi.org/10.1111/j.1530-0277.1996.tb01951.x>.
- Mattson, S. N., E. P. Riley, T. L. Jernigan, et al. 1994. "A Decrease in the Size of the Basal Ganglia Following Prenatal Alcohol Exposure: A Preliminary Report." *Neurotoxicology and Teratology* 16, no. 3: 283–289. [https://doi.org/10.1016/0892-0362\(94\)90050-7](https://doi.org/10.1016/0892-0362(94)90050-7).
- Næss-Schmidt, E., A. Tietze, J. U. Blicher, et al. 2016. "Automatic Thalamus and Hippocampus Segmentation From MP2RAGE: Comparison of Publicly Available Methods and Implications for DTI Quantification." *International Journal of Computer Assisted Radiology and Surgery* 11, no. 11: 1979–1991. <https://doi.org/10.1007/s11548-016-1433-0>.
- Nardelli, A., C. Lebel, C. Rasmussen, G. Andrew, and C. Beaulieu. 2011. "Extensive Deep Gray Matter Volume Reductions in Children and Adolescents With Fetal Alcohol Spectrum Disorders." *Alcoholism, Clinical and Experimental Research* 35, no. 8: 1404–1417. <https://doi.org/10.1111/j.1530-0277.2011.01476.x>.
- Neto, L. L., E. Oliveira, F. Correia, and A. G. Ferreira. 2008. "The Human Nucleus Accumbens: Where Is It? A Stereotactic, Anatomical and Magnetic Resonance Imaging Study." *Neuromodulation: Technology at the Neural Interface* 11, no. 1: 13–22. <https://doi.org/10.1111/j.1525-1403.2007.00138.x>.
- Norman, L. J., C. Carlisi, S. Lukito, et al. 2016. "Structural and Functional Brain Abnormalities in Attention-Deficit/Hyperactivity Disorder and Obsessive-Compulsive Disorder: A Comparative Meta-Analysis." *JAMA Psychiatry* 73, no. 8: 815–825. <https://doi.org/10.1001/jamapsychiatry.2016.0700>.

- Patenaude, B., S. M. Smith, D. N. Kennedy, and M. Jenkinson. 2011. "A Bayesian Model of Shape and Appearance for Subcortical Brain Segmentation." *NeuroImage* 56, no. 3: 907–922. <https://doi.org/10.1016/j.neuroimage.2011.02.046>.
- Pedregosa, F., G. Varoquaux, A. Gramfort, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12, no. 85: 2825–2830.
- Popova, S., S. Lange, K. Shield, et al. 2016. "Comorbidity of Fetal Alcohol Spectrum Disorder: A Systematic Review and Meta-Analysis." *Lancet* 387, no. 10022: 978–987. [https://doi.org/10.1016/S0140-6736\(15\)01345-8](https://doi.org/10.1016/S0140-6736(15)01345-8).
- Puonti, O., J. E. Iglesias, and K. Van Leemput. 2016. "Fast and Sequence-Adaptive Whole-Brain Segmentation Using Parametric Bayesian Modeling." *NeuroImage* 143: 235–249. <https://doi.org/10.1016/j.neuroimage.2016.09.011>.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing* [Logiciel]. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rakić, M., M. Cabezas, K. Kushibar, A. Oliver, and X. Lladó. 2020. "Improving the Detection of Autism Spectrum Disorder by Combining Structural and Functional MRI Information." *NeuroImage: Clinical* 25: 102181. <https://doi.org/10.1016/j.nicl.2020.102181>.
- Raznahan, A., P. W. Shaw, J. P. Lerch, et al. 2014. "Longitudinal Four-Dimensional Mapping of Subcortical Anatomy in Human Development." *Proceedings of the National Academy of Sciences* 111, no. 4: 1592–1597. <https://doi.org/10.1073/pnas.1316911111>.
- Reardon, P. K., L. Clasen, J. N. Giedd, et al. 2016. "An Allometric Analysis of Sex and Sex Chromosome Dosage Effects on Subcortical Anatomy in Humans." *Journal of Neuroscience* 36, no. 8: 2438–2448. <https://doi.org/10.1523/JNEUROSCI.3195-15.2016>.
- Riley, E. P., and C. L. McGee. 2005. "Fetal Alcohol Spectrum Disorders: An Overview With Emphasis on Changes in Brain and Behavior." *Experimental Biology and Medicine* 230, no. 6: 357–365. <https://doi.org/10.1177/15353702-0323006-03>.
- Rottschy, C., R. Langner, I. Dogan, et al. 2012. "Modelling Neural Correlates of Working Memory: A Coordinate-Based Meta-Analysis." *NeuroImage* 60, no. 1: 830–846. <https://doi.org/10.1016/j.neuroimage.2011.11.050>.
- Roussotte, F. F., K. K. Sulik, S. N. Mattson, et al. 2012. "Regional Brain Volume Reductions Relate to Facial Dysmorphology and Neurocognitive Function in Fetal Alcohol Spectrum Disorders." *Human Brain Mapping* 33, no. 4: 920–937. <https://doi.org/10.1002/hbm.21260>.
- Sankar, T., M. T. M. Park, T. Jawa, et al. 2017. "Your Algorithm Might Think the Hippocampus Grows in Alzheimer's Disease: Caveats of Longitudinal Automated Hippocampal Volumetry." *Human Brain Mapping* 38, no. 6: 2875–2896. <https://doi.org/10.1002/hbm.23559>.
- Schneider, M. L., C. F. Moore, and M. M. Adkins. 2011. "The Effects of Prenatal Alcohol Exposure on Behavior: Rodent and Primate Studies." *Neuropsychology Review* 21: 186–203. <https://doi.org/10.1007/s11065-011-9168-8>.
- Schulz, M.-A., B. T. T. Yeo, J. T. Vogelstein, et al. 2020. "Different Scaling of Linear Models and Deep Learning in UKBiobank Brain Images Versus Machine-Learning Datasets." *Nature Communications* 11, no. 1: 4238. <https://doi.org/10.1038/s41467-020-18037-z>.
- Schulz, M.-A., D. Bzdok, S. Haufe, J.-D. Haynes, and K. Ritter. 2024. "Performance Reserves in Brain-Imaging-Based Phenotype Prediction." *Cell Reports* 43, no. 1: 113597. <https://doi.org/10.1016/j.celrep.2023.113597>.
- Streissguth, A. P., and P. Dehaene. 1993. "Fetal Alcohol Syndrome in Twins of Alcoholic Mothers: Concordance of Diagnosis and IQ." *American Journal of Medical Genetics* 47, no. 6: 857–861. <https://doi.org/10.1002/ajmg.1320470612>.
- Sulik, K. K., J. M. Lauder, and D. B. Dehart. 1984. "Brain Malformations in Prenatal Mice Following Acute Maternal Ethanol Administration." *International Journal of Developmental Neuroscience* 2, no. 3: 203–214. [https://doi.org/10.1016/0736-5748\(84\)90014-5](https://doi.org/10.1016/0736-5748(84)90014-5).
- Sullivan, E. V., E. M. Moore, B. Lane, K. M. Pohl, E. P. Riley, and A. Pfefferbaum. 2020. "Graded Cerebellar Lobular Volume Deficits in Adolescents and Young Adults With Fetal Alcohol Spectrum Disorders (FASD)." *Cerebral Cortex* 30, no. 9: 4729–4746. <https://doi.org/10.1093/cercor/bhaa020>.
- Temple, V. K., J. L. Cook, K. Unsworth, H. Rajani, and M. Mela. 2019. "Mental Health and Affect Regulation Impairment in Fetal Alcohol Spectrum Disorder (FASD): Results From the Canadian National FASD Database." *Alcohol and Alcoholism* 54, no. 5: 545–550. <https://doi.org/10.1093/alcalc/agz049>.
- Thulin, M. 2021. *Modern bStatistics With R: From Wrangling and Exploring Data to Inference and Predictive Modelling*. Eos Chasma Press. <https://www.modernstatisticswithr.com/>.
- Toro, R., M. Chupin, L. Garnero, et al. 2009. "Brain Volumes and Val66Met Polymorphism of the BDNF Gene: Local or Global Effects?" *Brain Structure and Function* 213, no. 6: 501–509. <https://doi.org/10.1007/s00429-009-0203-y>.
- Tran, E. L., L. J. England, Y. Park, C. H. Denny, and S. Y. Kim. 2023. "Systematic Review: Polysubstance Prevalence Estimates Reported During Pregnancy, US, 2009–2020." *Maternal and Child Health Journal* 27, no. 3: 426–458. <https://doi.org/10.1007/s10995-023-03592-w>.
- Treit, S., Z. Chen, D. Zhou, et al. 2017. "Sexual Dimorphism of Volume Reduction but Not Cognitive Deficit in Fetal Alcohol Spectrum Disorders: A Combined Diffusion Tensor Imaging, Cortical Thickness and Brain Volume Study." *NeuroImage: Clinical* 15: 284–297. <https://doi.org/10.1016/j.nicl.2017.05.006>.
- Tremblay, L., Y. Worbe, S. Thobois, V. Sgambato-Faure, and J. Féger. 2015. "Selective Dysfunction of Basal Ganglia Subterritories: From Movement to Behavioral Disorders." *Movement Disorders* 30, no. 9: 1155–1170. <https://doi.org/10.1002/mds.26199>.
- Tsang, T. W., B. R. Lucas, H. Carmichael Olson, R. Z. Pinto, and E. J. Elliott. 2016. "Prenatal Alcohol Exposure, FASD, and Child Behavior: A Meta-Analysis." *Pediatrics* 137, no. 3: e20152542. <https://doi.org/10.1542/peds.2015-2542>.
- Varoquaux, G., and O. Colliot. 2023. "Evaluating Machine Learning Models and Their Diagnostic Value." In *Machine Learning for Brain Disorders*, edited by O. Colliot, 601–630. Springer US. https://doi.org/10.1007/978-1-0716-3195-9_20.
- Virtanen, P., R. Gommers, T. E. Oliphant, et al. 2020. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python." *Nature Methods* 17, no. 3: 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
- Wang, X., V. C. Cuzon Carlson, C. Studholme, et al. 2020. "In Utero MRI Identifies Consequences of Early-Gestation Alcohol Drinking on Fetal Brain Development in Rhesus Macaques." *Proceedings of the National Academy of Sciences* 117, no. 18: 10035–10044. <https://doi.org/10.1073/pnas.1919048117>.
- Weyrauch, D., M. Schwartz, B. Hart, M. G. Klug, and L. Burd. 2017. "Comorbid Mental Disorders in Fetal Alcohol Spectrum Disorders: A Systematic Review." *Journal of Developmental & Behavioral Pediatrics* 38, no. 4: 283–291. <https://doi.org/10.1097/DBP.0000000000000440>.
- Williams, C. M., H. Peyre, R. Toro, and F. Ramus. 2021. "Neuroanatomical Norms in the UK Biobank: The Impact of Allometric Scaling, Sex, and Age." *Human Brain Mapping* 42, no. 14: 4623–4642. <https://doi.org/10.1002/hbm.25572>.

Yushkevich, P. A., J. Piven, H. C. Hazlett, et al. 2006. “User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability.” *NeuroImage* 31, no. 3: 1116–1128. <https://doi.org/10.1016/j.neuroimage.2006.01.015>.

Zhuang, Q., L. Xu, F. Zhou, et al. 2021. “Segregating Domain-General From Emotional Context-Specific Inhibitory Control Systems—Ventral Striatum and Orbitofrontal Cortex Serve as Emotion-Cognition Integration Hubs.” *NeuroImage* 238: 118269. <https://doi.org/10.1016/j.neuroimage.2021.118269>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.