

Bias in HD-ISS staging introduced by the FreeSurfer cross-sectional stream: Insights from the Huntington's Disease Young Adult Study (HD-YAS)

Journal of Huntington's Disease

1–7

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/18796397251366900

journals.sagepub.com/home/hun

Harry Knights¹ , Annabelle Coleman¹ , Mena Farag¹ , Michela Leocadi¹ , Michael Murphy¹, Kate Fayer¹ , Olivia Thackeray¹, Douglas Langbehn² , Nicola Hobbs¹, Sarah J Tabrizi¹ , Rachael I Scahill¹ and the HD-YAS investigators

Abstract

Huntington's Disease Integrated Staging System (HD-ISS) stages are likely inclusion criteria in future clinical trials. Stage I volumetric cut-offs were derived using the FreeSurfer longitudinal stream (LG). However, trials will require cross-sectional stream (CS) application with one MRI. Volumetric outputs are not robust to software type or version. T1-weighted images from 88 participants with MRIs from baseline and follow-up HD-YAS visits were segmented using both streams. CS calculated smaller caudate and putamen volumes adjusted for total intracranial volume, with greater reduction for larger volumes, shifting towards HD-ISS stage I. CS-specific cut-offs need to be established before application to clinical trials.

Keywords

huntington's disease, FreeSurfer, MRI, segmentation, caudate, putamen, HD-ISS

Received: 19 December 2024; accepted: 8 July 2025

Introduction

Huntington's disease (HD) is a devastating progressive neurodegenerative disorder for which there remains no disease-modifying therapy. Inherited CAG repeat expansion undergoes further somatic expansion towards a critical length in vulnerable cell types, particularly the medium-spiny neurons in the striatum,¹ causing neuronal damage, dysfunction, and death.² This results in early atrophy within the striatum³ which extends over time to other sub-cortical and cortical regions.^{4,5}

The HD Young Adult Study (HD-YAS) is a deeply phenotyped far-from-onset observational study of closely matched HD gene expanded (HDGE) and control groups, assessed longitudinally at two visits (V1 and V2) approximately 4.5 years apart.³

The Huntington's Disease Integrated Staging System (HD-ISS) re-defined pre-manifest HD as stage 0 (without detectable biomarkers of pathophysiology) and 1 (caudate and/or putamen atrophy below the 5th percentile in healthy controls).⁶

We now enter a pivotal period for HD research, with multiple ongoing clinical trials of disease-modifying

therapies.⁷ Secondary prevention trials aim to target underlying disease pathobiology before significant neurodegeneration, with the goal of preserving the functional integrity of remaining striatal circuitry. Certain experimental therapies are also aiming to interfere with very early pathobiological mechanisms, including somatic expansion.^{8,9} HD-ISS stages 0 and 1 are therefore possible inclusion criteria in future secondary prevention trials.

The HD-ISS used the FreeSurfer (FS) longitudinal stream (LG) to define stage 1 cut-offs, with the aim of reducing within-subject variability.¹⁰ However, clinical trials will

¹Huntington's Disease Centre, Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, University College London, London, UK

²Department of Psychiatry and Biostatistics, Carver College of Medicine and College of Public Health, University of Iowa, Iowa City, IA, USA

Corresponding author:

Harry Knights, Huntington's Disease Centre, Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, University College London, London, UK.

Email: skgthkn@ucl.ac.uk

likely require application to participants with a single MRI, necessitating cross-sectional stream (CS) processing. MRI-derived volumetric outputs are not robust to pipeline¹¹ or software version,¹² and the impact of stream on volumes and staging remains unknown.

This study will explore the relationships and differences between caudate, putamen, and total intracranial volumes (TICV) derived from FS CS and LG in HD-YAS.

Methods

Study participants

HD-YAS participants were originally recruited from Enroll-HD (<https://www.enroll-hd.org/>), regional genetic and HD centers, the Huntington's Disease Association (<https://www.hda.org.uk/>), and the Huntington's Disease Youth Organisation (<https://hdyo.org/>). Participants were aged 18–40 inclusive and were excluded if they had a history of drug and/or alcohol abuse, significant co-morbidity, or contraindications to MRI. The HDGE group were required to have no clinical diagnostic motor features of HD (Unified Huntington's Disease Rating Scale [UHDRS] Diagnostic Confidence Level < 4), CAG repeat expansion length ≥ 40 , and Disease Burden Scores (DBS)¹³ ≤ 240 . Controls were at-risk gene-negative family members (CAG < 36), genetically unrelated family members, and members of the wider HD community. HDGE and control groups were matched for age, sex, and education using means and variances. Participants were enrolled from August 2017 to April 2019 for visit 1 (V1), and April 2022 to January 2024 for visit 2 (V2). 88 participants (54 HDGE individuals and 34 controls) with neuroimaging performed at both V1 and V2 were included in this study.

MRI acquisition

All MRIs were acquired using the same research-dedicated 3-Tesla Prisma scanner (Siemens Healthcare, Erlangen, Germany). T1-weighted images were acquired using a 3D Magnetization Prepared Rapid Gradient Echo (MPRAGE) with the following parameters: repetition time = 2530 ms; time to echo = 3.34 ms; inversion time = 1100 ms; flip angle = 7°; field of view = $256 \times 256 \times 176$ mm³; and resolution = $1.0 \times 1.0 \times 1.0$ mm³.

HD-YAS MRI scans are high quality and well-standardized due to the use of the same research dedicated MRI scanner, imaging protocols optimized for grey/white segmentation, experienced radiographers, and minimal motion artefacts in the far-from-onset HDGE group.

FreeSurfer

All segmentations were run on FS version 6.0.1 to mirror the methodology used to define staging cut-offs⁶ and

because segmentations are not robust to software version due to modifications to segmentation algorithms.¹² All segmentations were run on the same operating system, since this has also been shown to influence segmentation outputs.^{14,15} The explanation for this is complex and arises from a combination of factors, including how different operating systems handle floating-point numbers (i.e., rounding errors), how they order calculations, and how FS is compiled for each system (how it is integrated with the operating system).¹⁵

For CS, segmentations were generated using recon-all.^{16–18} In brief, T1-weighted images undergo skull-stripping, an affine transform to MNI305 space,¹⁹ intensity homogenization, and a non-linear transform. Probability distributions for voxel location and intensity are derived from the Talairach atlas.²⁰ Volume is calculated as the number of voxels of known size (usually 1 mm³) within the region-of-interest. TICV is inferred from the scaling factor required for the affine transformation to Talairach space.²¹

For LG, CS segmentations from V1 and V2 were combined to create an unbiased within-subject template using recon-base.^{22,23} Image processing is then initialized using common information from the within-subject template (avoiding interpolation asymmetry), and each time-point is processed individually (without temporal regularization) using the recon-long command.¹⁰ LG applies a fixed affine transformation across time-points, therefore deriving the same TICV at all time-points.

A flow diagram containing the processing steps for both streams is displayed in Supplemental Figure 1.

Quality control

Volumes may remain within the normal range despite inaccurate segmentations and therefore segmentations must be reviewed visually.²⁴ CS and LG segmentations for V1 and V2 were quality controlled by a single investigator (HK) blinded to disease status and volume. Segmentations were considered to be 'pass' or 'fail' based on whether the segmented boundaries were significantly outside of the visible boundary. No segmentations were identified as gross failures for either stream. Manual editing of segmentations was not performed. TICV could not be quality controlled since it is based on the affine transform to the Talairach atlas and no region is generated, consistent with HD-ISS methodology.⁶

Statistical analyses

Differences between volumes were explored using Bland-Altman analysis.²⁵ A scatter plot was created in which the y-axis shows the difference between two volumes (A–B), and the x-axis shows the mean between two volumes ($[A + B]/2$). Systematic bias was described

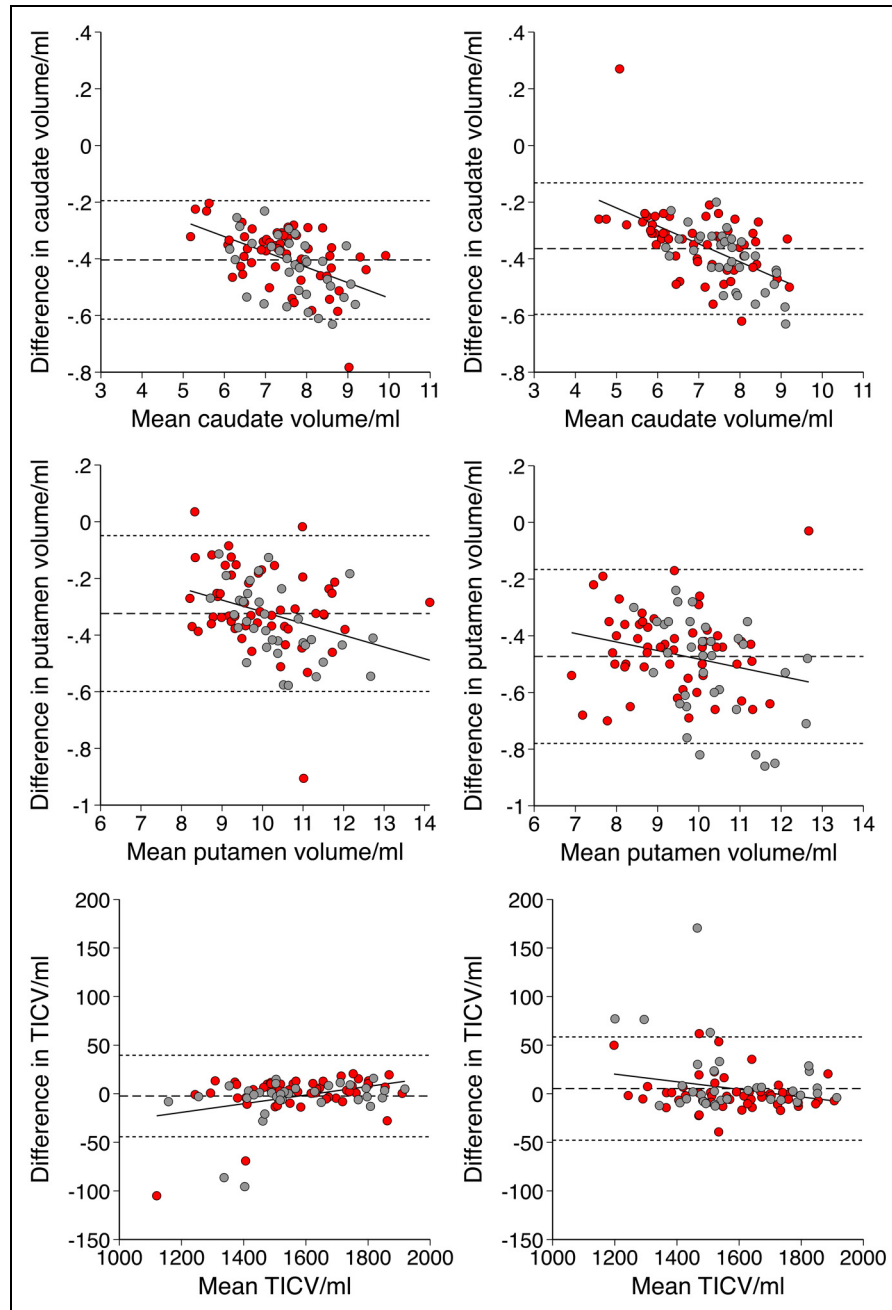


Figure 1. Bland-Altman plots with differences represented as cross-sectional (CS) – longitudinal (LG) stream volumes Total intracranial volume (TICV). Adjusted caudate and putamen are expressed as (raw volume/TICV)*1000. Left-sided plots are for visit 1 (V1) and right-sided plots are for visit 2 (V2). Data points are labelled as the HDGE (red) and control (grey) groups. Linear regression (solid), mean difference (dash), 95% limits of agreement (dots). Linear regression equations are depicted in Supplemental Table 3. (continued)

using the mean difference and 95% limits of agreement. Proportional bias was explored through linear regression analysis.²⁶ Similarities between CS and LG volumes were also assessed using intraclass correlation coefficient (ICC). A two-tailed p-value below 0.05 was considered statistically significant. All statistics were performed using Stata v17.0.

Results

Baseline demographics for participants are displayed in Supplemental Table 1.

All ICC values between CS and LG volumes were > 0.97 and were highly significant at $p < 0.0005$ (Supplemental Table 2).

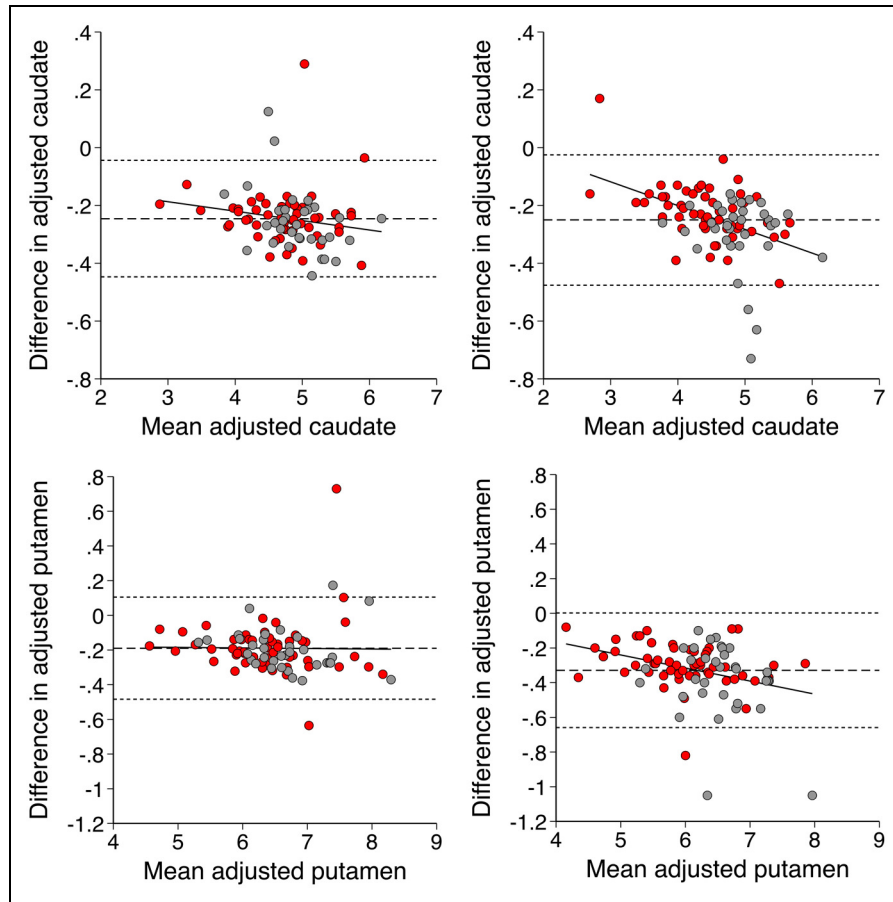


Figure 1. Continued.

Differences between volumes were described using Bland-Altman analysis (plots displayed in Figure 1 and systematic differences displayed in Supplemental Table 3). Compared to LG, CS calculated: i) smaller raw caudate (-5.5% V1, -5.1% V2, both $p < 0.00005$) and putamen volumes (-3.2% V1, -5.0% V2, both $p < 0.00005$), with a bias towards greater reduction for larger volumes and ii) similar TICV (-0.2% V1, $+0.3\%$ V2, both $p > 0.05$). This resulted in smaller adjusted caudate (-5.3% V1, -5.6% V2, both $p < 0.00005$, without proportional bias) and putamen (-3.0% V1, -5.5% V2, both $p < 0.00005$, with greater reduction for larger volumes).

These volumetric differences impacted staging, with CS shifting the HDGE group from stage 0 to 1, moving from 9/54 to 15/54 at V1 and 19/54 to 25/54 at V2 (Figure 2).

These findings suggest that CS estimates larger adjusted caudate and putamen volumes which are less likely to reach the stage 1 threshold, according to the existing longitudinally derived cut-offs.

Discussion

This study adds to the growing body of evidence describing the meaningful impact of subtle pipeline differences on

striatal segmentations, volumetric outputs, and HD-ISS staging in HDGE individuals. The use of HD-ISS staging as inclusion criteria in interventional trials will likely require application to HDGE individuals with a single MRI brain scan, precluding the use of LG.

Previous studies comparing CS and LG caudate and putamen segmentations are uncommon and have focused on reliability, showing improved test-retest reproducibility.^{10,27,28} This is useful for assessing sensitivity to detect subtle change over time, which impacts sample sizes and follow-up periods in interventional trials. However, the impact of stream of staging requires an exploration of measurement bias.

This study has shown that caudate, putamen, and total intracranial volumes varied greatly between CS and LG. The combined effect was that CS estimated smaller adjusted caudate and putamen volumes, with greater effect on larger volumes. This shifted the HDGE group towards stage 1 (Figure 2). In particular, the use of CS at V1, and LG at V2, as might be assumed to be a reasonable methodology to maximize the accuracy of individual segmentations with growing available data, functioned to substantially reduce the staging progression between time-points.

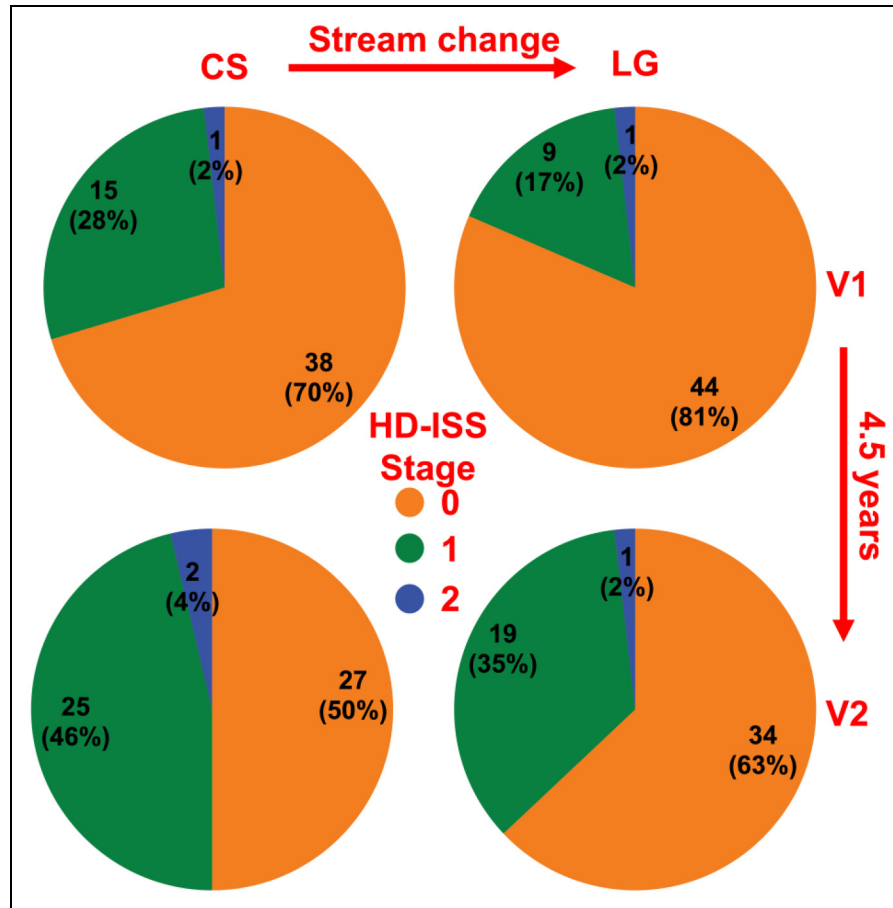


Figure 2. HD-ISS staging according to the FreeSurfer cross-sectional (CS) and longitudinal (LG) streams. Staging for the HDGE group ($n = 54$) with brain MRIs at visit 1 (V1) and visit 2 (V2) are displayed. Moving from left to right changes from CS to LG. Moving from top to bottom changes from V1 to V2.

It should be noted, however, that LG may be less suitable for estimating volumes when there is substantial atrophy over time. This limitation arises from a combination of factors: template bias (where a later atrophic scan can distort the segmentation of an earlier healthy scan); registration errors; and non-linear degeneration.¹⁰ However, with minimal atrophy during the transition from HD-ISS stage 0 and 1, LG remains appropriate for use in this context.

The mechanistic explanation for these differences is challenging to define. LS applies the same affine transform to all time-points, deriving the same TICV, which is approximately the average of the two cross-sectional TICVs (Supplemental Figure 2), explaining the non-significant difference in TICV. Differences in caudate and putamen segmentation likely relate to the creation of a within-subject median template image to initialize the segmentation.

Overall, applying the HD-ISS to interventional trials requires staging with CS. This calculates larger adjusted caudate and putamen volumes than LG, which was used to generate stage 1 cut-offs, shifting from stage 0 to 1 across time-points. The HD-ISS must urgently define

cut-offs derived from CS and make them publicly available before widespread application in clinical trials.

Acknowledgements

We would like to thank all the HD-YAS participants for their invaluable contribution.

We would also like to thank the HD-YAS investigators: Paul Zeun, Katherine Osborne-Crowley, Eileanoir B Johnson, Sarah Gregory, Christopher Parker, Jessica Lowe, Akshay Nair, Marina Papoutsis, Peter McColgan, Carlos Estevez-Fraga, Kate Fayer, Henny Wellington, Filipe B Rodrigues, Lauren M Byrne, Amanda Heslegrave, Harpreet Hyare, Hui Zhang, Edward J Wild, Geraint Rees (University College London); Claire O'Callaghan, Christelle Langley, Trevor W Robbins, Barbara J Sahakian (Cambridge University); and Douglas Langbehn (University of Iowa).

ORCID iDs

Harry Knights  <https://orcid.org/0000-0002-8184-7208>

Annabelle Coleman  <https://orcid.org/0000-0002-9045-0034>

Mena Farag  <https://orcid.org/0000-0002-0679-0117>
 Michela Leocadi  <https://orcid.org/0000-0001-7321-6054>
 Kate Fayer  <https://orcid.org/0000-0002-4198-2632>
 Douglas Langbehn  <https://orcid.org/0000-0002-7803-4679>
 Sarah J Tabrizi  <https://orcid.org/0000-0003-2716-2045>
 Rachael I Scahill  <https://orcid.org/0000-0002-1832-106X>

Ethical considerations

HD-YAS was approved by the London-Bloomsbury Research Ethics Committee (22/LO/0058).

Consent to participate

All participants provided written informed consent before enrolment.

Funding

HK is supported by the National Institute of Health Research Academic Clinical Fellowship. HD-YAS is funded by the Wellcome Trust (grant codes 200181/Z/15/Z and 223082/Z/21/Z – SJT is PI and RIS and NZH receive funding from this grant). We also thank the Wellcome Trust Centre for Human Neuroimaging (London, UK) who acquired the MRI scans. Part of this work and funding for SJT was supported by the UK Dementia Research Institute (DRI London UK) which receives its funding from DRI Ltd, funded by the UK Medical Research Council. Some of this work was undertaken at the University College London Hospital/University College London (London UK) supported by the UK's Department of Health National Institute of Health Biomedical Research Centre (London UK).

Declaration of conflicting interests

SJT is an Associate Editor of the Journal of Huntington's Disease, but was not involved in the peer-review process and did not have access to any information regarding peer-review.

Data availability statement

The data supporting the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy, ethical restrictions, or other concerns.

Supplemental material

Supplemental material for this article is available online.

References

1. Handsaker RE, Kashin S, Reed NM, et al. Long somatic DNA-repeat expansion drives neurodegeneration in huntington's disease. *Cell* 2025; 188: 623–639.e19.
2. Hong EP, MacDonald ME, Wheeler VC, et al. Huntington's disease pathogenesis: two sequential components. *J Huntingtons Dis* 2021; 10: 35.
3. Scahill RI, Zeun P, Osborne-Crowley K, et al. Biological and clinical characteristics of gene carriers far from predicted onset in the huntington's disease young adult study (HD-YAS): a cross-sectional analysis. *Lancet Neurol* 2020; 19: 502–512.
4. Tabrizi SJ, Langbehn DR, Leavitt BR, et al. Biological and clinical manifestations of huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *Lancet Neurol* 2009; 8: 791–801.
5. Tabrizi SJ, Scahill RI, Owen G, et al. Predictors of phenotypic progression and disease onset in premanifest and early-stage huntington's disease in the TRACK-HD study: analysis of 36-month observational data. *Lancet Neurol* 2013; 12: 637–649.
6. Tabrizi SJ, Schobel S, Gantman EC, et al. A biological classification of huntington's disease: the integrated staging system. *Lancet Neurol* 2022; 21: 632–644.
7. Estevez-Fraga C, Tabrizi SJ and Wild EJ. Huntington's disease clinical trials corner: march 2024. *J Huntingtons Dis* 2024; 13: 1–14.
8. Kennedy L, Evans E, Chen CM, et al. Dramatic tissue-specific mutation length increases are an early molecular event in huntington disease pathogenesis. *Hum Mol Genet* 2003; 12: 3359–3367.
9. O'Reilly D, Belgrad J, Ferguson C, et al. Di-valent siRNA-mediated silencing of MSH3 blocks somatic repeat expansion in mouse models of huntington's disease. *Mol Ther* 2023; 31: 1661–1674.
10. Reuter M, Schmansky NJ, Rosas HD, et al. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* 2012; 61: 1402–1418.
11. Mansoor NM, Vanniyasingam T, Malone I, et al. Validating automated segmentation tools in the assessment of caudate atrophy in huntington's disease. *Front Neurol* 2021; 12: 616272.
12. Knights H, Coleman A, Hobbs NZ, et al. Freesurfer software update significantly impacts striatal volumes in the huntington's disease young adult study and will influence HD-ISS staging. *J Huntingtons Dis* 2024; 13: 77–90.
13. Penney JB, Vonsattel JP, MacDonald ME, et al. CAG Repeat number governs the development rate of pathology in huntington's disease. *Ann Neurol* 1997; 41: 689–692.
14. Gronenschild EHB, Habets P, Jacobs HIL, et al. The effects of FreeSurfer version, workstation type, and macintosh operating system version on anatomical volume and cortical thickness measurements. *PLoS One* 2012; 7: e38234.
15. Glatard T, Lewis LB, da Silva RF, et al. Reproducibility of neuroimaging analyses across operating systems. *Front Neuroinform* 2015; 9: 135293.
16. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 2002; 33: 341–355.
17. Fischl B, Salat DH, Van Der Kouwe AJW, et al. Sequence-independent segmentation of magnetic resonance images. *Neuroimage* 2004; 23: S69–S84.
18. Fischl B. Freesurfer. *Neuroimage* 2012; 62: 774–781.
19. Evans AC, Collins DL, Mills SR, et al. 3D statistical neuroanatomical models from 305 MRI volumes. 1993 *IEEE Conference Record Nuclear Science Symposium and*

- Medical Imaging Conference*, San Francisco, CA, USA, 1993, pp. 1813–1817.
20. Talairach J and Tournoux P. *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging*. Stuttgart; New York: G. Thieme; New York: Thieme Medical Publishers, 1988.
 21. Buckner RL, Head D, Parker J, et al. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage* 2004; 23: 724–738.
 22. Reuter M, Rosas HD and Fischl B. Highly accurate inverse consistent registration: a robust approach. *Neuroimage* 2010; 53: 1181–1196.
 23. Reuter M and Fischl B. Avoiding asymmetry-induced bias in longitudinal image processing. *Neuroimage* 2011; 57: 19–21.
 24. Johnson EB, Gregory S, Johnson HJ, et al. Recommendations for the use of automated gray matter segmentation tools: evidence from huntington's disease. *Front Neurol* 2017; 8: 300602.
 25. Altman DG and Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician* 1983; 32: 307.
 26. Ho KM. Using linear regression to assess dose-dependent bias on a bland-altman plot. *J Emerg Crit Care Med* 2018; 2: 68–68.
 27. Jovicich J, Marizzoni M, Sala-Llonch R, et al. Brain morphometry reproducibility in multi-center 3T MRI studies: a comparison of cross-sectional and longitudinal segmentations. *Neuroimage* 2013; 83: 472–484.
 28. Hedges EP, Dimitrov M, Zahid U, et al. Reliability of structural MRI measurements: the effects of scan session, head tilt, inter-scan interval, acquisition sequence, FreeSurfer version and processing stream. *Neuroimage* 2022; 246: 118751.