Published in final edited form as:

Mov Disord. 2014 September 15; 29(11): 1429–1433. doi:10.1002/mds.26013.

MRI Striatal Volumes – A Biomarker for Clinical Trials in HD

Elizabeth H. Aylward, Ph.D.

Seattle Children's Research Institute Seattle, Washington, USA

Abstract

There is an abundance of research showing that MRI striatal volumes decrease long before diagnosis of HD and closely track disease progression. Additional research indicates that these volumetric measures meet important criteria for a biomarker that can be used in clinical trials: They are (1) objectively measureable; (2) able to predict known endpoints; and (3) associated with known mechanisms of pathology of the disease. It is recommended that HD researchers consider formal application to regulatory agencies for biomarker status of volumetric MRI striatal measures, as it is anticipated that these measures can contribute significantly in the assessment of treatment effectiveness in HD.

A biomarker (also known as a biological marker) is “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” [1]. This definition is generally equivalent to Mildvan's definition of a “natural history” biomarker [2], a “marker of disease severity that reflects underlying pathogenesis and predicts clinical events in the absence of treatment, thus establishing the biological plausibility of the marker.” Based on these two definitions, potential biomarkers can be evaluated for utility according to three criteria [3]: They must be (1) objectively measureable; (2) able to predict known endpoints; and (3) associated with known mechanisms of pathology of the disease.

As will be demonstrated, volumetric striatal measures from MRI meet these three criteria, as well as other desirable qualities of a biomarker for clinical trials. Although evidence exists that supports the use of volumetric measures of other brain regions as potential outcome measures [4], the greatest amount of research and the most consistent results across studies support the use of striatal volumes, with measurement of caudate, putamen, and total striatum generally showing equally strong potential. Other neuroimaging methods, including MRS, DTI, and fMRI, have been studied as well, but to date findings have either been inconsistent or insufficient to justify their use as outcome measures. Further study may, of course, demonstrate better utility for these methods than for volumetric striatal measures, especially as it is possible that these measures might be able to detect change earlier or faster than structural MRI, and may be sensitive to reversal of the disease process, which is unexpected for structural volumes.

Objectively measured

Clinical trials in HD will generally require samples too large to be recruited from a single site. For many biomarkers, such as the Total Motor Scale of the Unified Huntington's Disease Rating Scale (UHDRS) [5], careful selection and training is required to achieve

reliability among raters [6], and ceiling and floor effects can hamper the detection of changes in pre-HD and advanced patients [7]. Reliability across sites and across time is somewhat more easily achievable for structural MRI measures, assuming that identical scan equipment (hardware and software) and scan protocols can be used, and that volumetric measurement is done using a standard method at a single site. Previous studies in other patient populations have analyzed data from multiple scanners and evaluated the validity of combining multi-scanner datasets (e.g., [8–11]). In general, these studies have demonstrated greater between-group differences than between-scanner differences and have concluded that the effect of multi-site scanning does not confound study results as long as cases and controls are equally represented at each site. If it is not possible to use identical scan equipment across sites, methods have been developed to create MRI sequences that yield nearly-equivalent results across scanners from different vendors. Assessing reliability across sites has involved use of MRI phantoms and human volunteers who are scanned at each site (e.g., [12–14]). Based on these studies, investigators have developed methods for image correction that reduce some of the inevitable inter-scanner differences. Similarly, reliability across time is essential for clinical trials, and methods have been developed to assess stability in MRI acquisition over time within a single site (e.g., [15]) and provide methods for statistical correction, when necessary, for changes in scanner software.

Reliability for measurements can be achieved by using programs that automatically segment brain structures, including atlas-based methods (e.g., [16–17]), voxel-based morphometry with statistical parametric mapping [18], deformation tensor morphometry (e.g., [19–20]) and boundary shift integral methods (e.g., [21–22]), but validity is sometimes an issue with these methods [23–25]. For example, studies using different automated methods have yielded contradictory results regarding the relative longitudinal change in gray vs. white matter volume in pre-HD [4,26], and studies of anterior cingulate have shown increased volume in HD using Freesurfer [27], but decreased volume using voxel-based morphometry [28, 29] and hand-traced methods [30]. Such programs must, therefore, be used cautiously and be validated on a sample representative of the subjects for whom the clinical trial is designed. Rater involvement at each step of the processing pipeline will insure that scans of limited quality are discarded and will verify that automatic segmentation results validly reflect actual structure volumes. A recent improvement in Freesurfer, based on analysis of longitudinal data is reported to yield better reliability than analyses based on cross-sectional data [31–32].

Predict known endpoints

The clinical endpoint of most interest in preclinical HD is onset of diagnosable signs, and for clinical trials in this population the biomarker must, therefore, be associated in a predictable manner with time to diagnosis. Clinical endpoints for HD patients already diagnosed will include standardized measures of motor, cognitive, psychiatric, and functional impairments, as well as overall survival. The type of clinical endpoint will, of course, be determined by the expected impact of the treatment. Neuroimaging measures associated with these signs and symptoms may be useful if they are tightly associated with the clinical outcomes. As discussed above, neuroimaging methods may yield more reliable measures than clinical outcomes, such as diagnosis of HD, UHDRS Total Motor Score, or

functional capacity, especially when these clinical measures are performed by multiple raters. If neuroimaging measures are demonstrated to be highly correlated with these clinical measures (most reliably determined in single-site studies), they may serve as more consistent indicators of clinical progression, especially when multiple sites are required for a trial. More reliable measures will yield less measurement error and will thus reduce the number of subjects required to demonstrate a treatment effect. Use of neuroimaging measures will be particularly applicable in pre-HD, where current measures of clinical progression are less sensitive to the earliest changes.

Data are available indicating that correlations between structural MRI measures and clinical measures are highly significant. In a single-site study of pre-HD, correlations between estimated years to onset (based on age, CAG repeat length, and parent's age at onset) and striatal volume were $r = 0.76$ ($p < 0.001$) for putamen and $r = 0.69$ ($p = 0.002$) for caudate [33]; correlations were somewhat lower in the large multi-site PREDICT-HD study ($r = .54$ for caudate and $.55$ for putamen; $p < .0001$) [34]. In symptomatic HD, significant correlations have been demonstrated between striatal volume and disease burden (based on CAG and age) [35]. In both TRACK-HD and PREDICT-HD, prospective multi-site studies, volume of striatum has been demonstrated to be one of the best predictors of disease onset in pre-HD [36, 37]. In both pre-HD and symptomatic individuals, very strong associations have been found between striatal volume and measures of both motor and cognitive impairment [38–42]. Striatal volumes are not, however, correlated with psychiatric symptoms in pre-HD or symptomatic HD [39, 43], perhaps due to the large variability in psychiatric impairments in HD and because of the effectiveness of psychiatric treatments. Thus, MRI striatal volumes will likely be poor outcome measures if the treatment is targeted toward psychiatric symptoms. Furthermore, striatal volumes are not correlated with functional measures in pre-HD subjects [39], most likely because of ceiling effects on current measures of functional capacity in this population.

Associated with known mechanisms of pathology of the disease

As discussed above, striatal volumes are highly associated with measures of disease burden and are predictive of disease onset. It is believed that striatal atrophy, which can be directly assessed by MRI, reflects neuronal death caused by build-up of neurotoxins. Striatal atrophy is the hallmark neuropathological feature of HD [44], and a recent neuropathological study [45] further indicates that putamen volume is related to number and volume of medium spiny neurons which, together with neuronal loss in subthalamic nuclei, contribute specifically to motor impairment in HD.

Other considerations

In addition to these primary qualities of a biomarker, other criteria are desirable for outcome measures that are practical and cost-effective for clinical trials. For example, a biomarker will generally be most useful for a clinical trial if, in the absence of treatment, the biomarker shows linear and rapid change. (It can be argued, of course, that some biomarkers may be relatively stable in the absence of treatment (e.g., brain metabolites assessed with magnetic resonance spectroscopy), but that treatment might be able to reverse deficits. However, this

claim cannot be assessed until an effective treatment is identified.) Striatal volume measures show decline that is both relatively rapid and linear once the atrophy begins, approximately 10–15 years prior to diagnosis. Data from PREDICT-HD found an annual decrease in striatal volume of 3.29 percent (from baseline volume of 11.43cc) for pre-HD subjects who were within 9 years of estimated onset, as compared to 0.6 percent annual decrease (from baseline of 16.48 cc) for control subjects [4]. When using absolute volumes (not ratios using baseline striatal volume as a correction factor), the rate of decline is relatively linear throughout the course of disease, starting approximately 15 years before estimated onset of motor symptoms [4].

It is also important to note that volumetric MRI measures are on a ratio scale, allowing a meaningful understanding that, for example, a striatum of 8cc is twice as large as one of 4cc. Few clinical measures have this characteristic, making it more difficult for longitudinal change to be linear over time. Finally, as compared with many clinical measures, striatal volumes show much less intra-subject variability over time [26]. Figure 1 shows longitudinal change in putamen volume for subjects in the PREDICT-HD study over a period of 11 years, as compared to scores on the UHDRS Total Motor Scale. Putamen volume change is more linear than TMS change, and the intra-individual variability over time is more consistent, suggesting less measurement error.

An additional advantage to striatal volume measures is that there are no ceiling effects and virtually no floor effects, at least in the population of subjects whose level of disease severity does not preclude scanning. Ceiling effects are often a problem for clinical measures (e.g., TMS and functional scales) for pre-HD subjects, especially those who are far from motor diagnosis. As suggested by the comments above, striatal volume measures may have particular utility in early pre-HD, unlike many clinical measures. Striatal volume is already decreased by approximately 7% in pre-HD subjects who are > 15 years from onset, and the rate of change is significantly faster than for controls. The rate of change during this far-from-onset stage (1.79% per year) is, however, significantly slower than the rate of change for those closer to onset (3.29% per year), and thus clinical trials in this population would require much larger samples than in studies using subjects who are somewhat farther along in the disease process [4]. Several studies have provided estimates of sample sizes that would be needed for studies using MRI striatal volume as an outcome measure in clinical trials at different disease stages (see [48] for a review). Finally, MRI assessment is fairly noninvasive and relatively inexpensive, especially if scanning is limited to a single 3-D structural image.

In summary, a large body of research over the past two decades supports the utility of volumetric MRI striatal measures as potential outcome measures in future clinical trials. It is recommended that HD researchers consider these measures be included, at least as secondary outcome measures, so that additional data can be collected to justify their use in future clinical trials. In the past, biomarkers have been accepted through ad hoc pathways at both the U.S. Food and Drug Administration (FDA) and the European Medicines Agency, on a case-by-case basis tailored to a particular drug of interest [49]. The FDA has recently established a more formalized Biomarker Qualification Program to accelerate the process by which new biomarkers are integrated in the development of therapies (<http://www.fda.gov/>

[Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm284076.htm](#)). Although no specific criteria are provided for approval of biomarkers, it is clear that much research has already been completed to address the validation and qualification process for structural MRI striatal measures. It is recommended that HD researchers consider formal application to regulatory agencies for biomarker status of volumetric MRI striatal measures, as it is anticipated that these measures can contribute significantly in the assessment of treatment effectiveness in HD.

References

1. Biomarkers Definition Workgroup. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther.* 2001; 69:89–95. [PubMed: 11240971]
2. Mildvan, D. In: Downing, GJ., editor. Biomarkers as surrogate endpoints in clinical trials for HIV/AIDS: a model for other diseases?; Biomarkers and Surrogate Endpoints: Clinical Research and Applications. Proceedings of NIH-FDA Conference; April 15–16, 1999; Bethesda, MD. Amsterdam: Elsevier; 2000. p. 13–17.
3. Aylward EH. Change in MRI striatal volumes as a biomarker in preclinical Huntington's disease. *Brain Res Bull.* 2007; 72:152–158. [PubMed: 17352939]
4. Aylward EH, Nopoulos PC, Ross CA, et al. Longitudinal change in regional brain volumes in prodromal Huntington disease. *J Neurol Neurosurg Psychiatry.* 2011; 82:405–410. [PubMed: 20884680]
5. Huntington Study Group. Unified Huntington's disease rating scale: reliability and consistency. *Mov Disord.* 1996; 11:136–142. [PubMed: 8684382]
6. Hogarth P, Kayson E, Kieburz K, et al. Interrater agreement in the assessment of motor manifestations of Huntington's disease. *Mov Disord.* 2005; 20:293–297. [PubMed: 15584032]
7. Youssov K, Dolbeau G, Maison P, et al. Unified Huntington's Disease Rating Scale for advanced patients: validation and follow-up study. *Mov Disord.* 2013; 28:1717–1723. [PubMed: 24166899]
8. Takao H, Hayashi N, Ohtomo K. Effects of study design in multi-scanner voxel-based morphometry studies. *Neuroimage.* 2014; 84:133–140. [PubMed: 23994315]
9. Fennema-Notestine C, Gamst AC, Quinn BT, et al. Feasibility of multi-site clinical structural neuroimaging studies of aging using legacy data. *Neuroinformatics.* 2007; 5:235–245. [PubMed: 17999200]
10. Pardoe H, Pell GS, Abbott DF, Berg AT, Jackson GD. Multi-site voxel-based morphometry: methods and a feasibility demonstration with childhood absence epilepsy. *Neuroimage.* 2008; 42:611–616. [PubMed: 18585930]
11. Stonnington CM, Tan G, Kloppel S, et al. Interpreting scan data acquired from multiple scanners: a study with Alzheimer's disease. *Neuroimage.* 2008; 39:1180–1185. [PubMed: 18032068]
12. Han X, Jovicich J, Salat D, et al. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage.* 2006; 32:180–194. [PubMed: 16651008]
13. Jovicich J, Czanner S, Greve D, et al. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage.* 2006; 30:436–443. [PubMed: 16300968]
14. Gouttard S, Styner M, Prastawa M, Piven J, Gerig G. Assessment of reliability of multi-site neuroimaging via traveling phantom study. *Med Image Comput Comput Assist Interv.* 2008; 11:263–270. [PubMed: 18982614]
15. Leow AD, Klunder AD, Jack CR Jr, et al. Longitudinal stability of MRI for mapping brain change using tensor-based morphometry. *Neuroimage.* 2006; 31:627–640. [PubMed: 16480900]
16. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis: Segmentation and surface reconstruction. *Neuroimage.* 1999; 9:179–194. [PubMed: 9931268]

17. Magnotta VA, Harris G, Andreasen NC, O'Leary DS, Yuh WT, Heckel D. Structural MR image processing using the BRAINS2 toolbox. *Comput Med Imaging Graph.* 2002; 26:251–264. [PubMed: 12074920]
18. Ashburner J, Friston KJ. Voxel-based morphometry--the methods. *Neuroimage.* 2000; 11:805–821. [PubMed: 10860804]
19. Miller MI, Beg MF, Ceritoglu C, Stark C. Increasing the power of functional maps of the medial temporal lobe by using large deformation diffeomorphic metric mapping. *Proc Natl Acad Sci USA.* 2005; 102:9685–9690. [PubMed: 15980148]
20. Studholme C, Cardenas V, Blumenfeld R, et al. Deformation tensor morphometry of semantic dementia with quantitative validation. *Neuroimage.* 2004; 21:1387–1398. [PubMed: 15050564]
21. Fox NC, Freeborough PA. Brain atrophy progression measured from registered serial MRI: validation and application to Alzheimer's disease. *J Magn Reson Imaging.* 1997; 7:1069–1075. [PubMed: 9400851]
22. Smith SM, Zhang Y, Jenkinson M, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage.* 2002; 17:479–489. [PubMed: 12482100]
23. Dewey J, Hana G, Russell T, Price J, et al. Reliability and validity of MRI-based automated volumetry software relative to auto-assisted manual measurement of subcortical structures in HIV-infected patients from a multisite study. *Neuroimage.* 2010; 51:1334–1344. [PubMed: 20338250]
24. Henley SM, Ridgway GR, Scahill RI, et al. Pitfalls in the use of voxel-based morphometry as a biomarker: examples from Huntington disease. *AJNR Am J Neuroradiol.* 2010; 31:711–719. [PubMed: 20037137]
25. Wenger E, Mårtensson J, Noack H, et al. Comparing manual and automatic segmentation of hippocampal volumes: Reliability and validity issues in younger and older brains. *Hum Brain Mapp.* 2014 Feb 14. [Epub ahead of print]. 10.1002/hbm.22473
26. Paulsen JS, Long JD, Johnson HJ, et al. Clinical and biomarker changes in premanifest Huntington disease show trial feasibility: A decade of the PREDICT-HD study. *Front Aging Neurosci.* 2014; 6:78. [PubMed: 24795630]
27. Rosas HD, Salat DH, Lee SY, et al. Cerebral cortex and the clinical expression of Huntington's disease: complexity and heterogeneity. *Brain.* 2008; 131:1057–1068. [PubMed: 18337273]
28. Tabrizi SJ, Langbehn DR, Leavitt BR, et al. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACKHD study: cross-sectional analysis of baseline data. *Lancet Neurol.* 2009; 8:791–801. [PubMed: 19646924]
29. Henley SM, Wild EJ, Hobbs NZ, et al. Relationship between CAG repeat length and brain volume in premanifest and early Huntington's disease. *J Neurol.* 2009; 256:203–212. [PubMed: 19266143]
30. Hobbs NZ, Pedrick AV, Say MJ, et al. The structural involvement of the cingulate cortex in premanifest and early Huntington's disease. *Mov Disord.* 2011; 26:1684–1690. [PubMed: 21557312]
31. Reuter M, Schmansky NJ, Rosas HD, Fischl B. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage.* 2012; 61:1402–1418. [PubMed: 22430496]
32. 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. [PubMed: 23668971]
33. Aylward EH, Sparks BF, Field KM, et al. Onset and rate of striatal atrophy in preclinical Huntington disease. *Neurology.* 2004; 63:66–72. [PubMed: 15249612]
34. Paulsen JS, Nopoulos PC, Aylward E, et al. Striatal and white matter predictors of estimated diagnosis for Huntington disease. *Brain Res Bull.* 2010; 82:201–207. [PubMed: 20385209]
35. Tabrizi SJ, Scahill RI, Durr A, et al. Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: The 12-month longitudinal analysis. *Lancet Neurol.* 2011; 10:31–42. [PubMed: 21130037]
36. Paulsen JS, Long JD, Johnson HJ, et al. Improving prediction of Huntington disease onset with clinical and imaging measures: A 10-year prospective study of converters. Under Review
37. Tabrizi SJ, Reilmann R, Roos RA, et al. Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: Analysis of 24 month observational data. *Lancet Neurol.* 2012; 11:42–53. [PubMed: 22137354]

38. Aylward EH, Brandt J, Codori AM, Mangus RS, Barta PE, Harris GJ. Reduced basal ganglia volume associated with the gene for Huntington's disease in asymptomatic at-risk persons. *Neurology*. 1994; 44:823–828. [PubMed: 8190282]
39. Aylward EA, Harrington DL, Mills JA, et al. Regional atrophy associated with cognitive and motor function in prodromal Huntington Disease. *J Huntington's Dis*. 2013; 2:477–489. [PubMed: 25062732]
40. Van den Bogaard SJA, Dumas EM, et al. Early atrophy of pallidum and accumbens nucleus in Huntington's disease. *J Neurol*. 2011; 258:412–420. [PubMed: 20936300]
41. Ruocco HH, Bonilha L, Li LM, Lopes-Cendes I, Cendes F. Longitudinal analysis of regional grey matter loss in Huntington disease: effects of the length of the expanded CAG repeat. *J Neurol Neurosurg Psychiatry*. 2008; 79:130–135. [PubMed: 17615168]
42. Ginestron A, Battaglini M, Diciotti S, et al. Magnetization transfer MR imaging demonstrates degeneration of the subcortical and cortical gray matter in Huntington disease. *Am J Neuroradiol*. 2010; 31:1807–1812. [PubMed: 20813872]
43. Scahill RI, Hobbs NZ, Say MJ, et al. Clinical impairment in premanifest and early Huntington's disease is associated with regionally specific atrophy. *Human Brain Mapping*. 2013; 34:519–529. [PubMed: 22102212]
44. Vonsattel JP, Myers RH, Stevens TJ, Ferrante RJ, Bird ED, Richardson EP. Neuropathological classification of Huntington's disease. *J Neuropathol Exp Neurol*. 1985; 44:559–577. [PubMed: 2932539]
45. Guo Z, Rudow G, Pletnikova O, et al. Striatal neuronal loss correlates with clinical motor impairment in Huntington's disease. *Mov Disord*. 2012; 27(11):1379–1386. [PubMed: 22975850]
46. Zhang Y, Long JD, Mills JA, Warner JH, Lu WJ, Paulsen S. PREDICT-HD Investigators and Coordinators of the Huntington Study Group. Indexing disease progression at study entry with individuals at-risk for Huntington disease. *Am J Med Genet B Neuropsychiatr Genet*. 2011; 156B: 751–763.
47. Ross CA, Aylward EH, Wild EJ, et al. Huntington disease: natural history, biomarkers and prospects for therapeutics. *Nat Rev Neurol*. 2014; 10:204–216. [PubMed: 24614516]
48. Georgiou-Karistianis N, Scahill R, Tabrizi SJ, Squitieri F, Aylward E. Structural MRI in Huntington's disease and recommendations for its potential use in clinical trials. *Neurosci Biobehav Rev*. 2013; 37:480–490. [PubMed: 23376047]
49. Goodsaid FM, Mendrick DL. Translational medicine and the value of biomarker qualification. *Sci Transl Med*. 2010; 2 (47):47ps44.

FIGURE 1. Comparison of Longitudinal Change in MRI Putamen Volume Measures and Total Motor Score in Pre-HD

Longitudinal change for individual subjects (thin lines) in the PREDICT-HD study for putamen volume (N = 1206) and UHDRS Total Motor Score (N = 1308). Thick lines are fitted spline curves and dashed lines indicate individuals who were diagnosed during the course of the study. Data collection occurred yearly for TMS and every other year for putamen volume, for up to 11 years per subject (median years in study = 6). CAP_D is a dynamic (time-varying) version of the CAP_E score [46] computed with current age (rather than age at entry), and can be interpreted as a type of CAG-adjusted age metric [47]. $CAP_E = (\text{age at entry}) * (CAG - 33.66)$, and is considered to be “disease burden” score that presumably indexes the cumulative toxicity of mutant huntingtin. Vertical line indicates average CAP_D at which diagnosis was made for this sample. Figure reproduced with permission from Paulsen et al., 2014.[26]