BRIEF REPORT

Age of Onset of Huntington's Disease in Carriers of Reduced Penetrance Alleles

Erin I. McDonnell,[1*] Yuanjia Wang,[1,2] Jill Goldman,[3] and Karen Marder[2,3,4,5]

[1]*Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, New York, USA* [2]*Department of Psychiatry, Columbia University Medical Center, New York, New York, USA* [3]*The Taub Institute for Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, New York, USA* [4]*Gertrude H. Sergievsky Center, Columbia University Medical Center, New York, New York, USA* [5]*Department of Neurology, Columbia University Medical Center, New York, New York, USA*

ABSTRACT: Background: Age of manifest Huntington's disease (HD) onset correlates with number of CAG repeats in the huntingtin gene. Little is known about onset with 36 to 39 repeats, the "reduced penetrance" (RP) range.

Objectives: We provide allele-specific estimates of HD penetrance (diagnostic confidence level of 4) for RP allele carriers.

Methods: We analyzed 431 pre-manifest RP allele carriers from Enroll-HD, the largest prospective observational HD study. Cumulative penetrance (CP) was estimated from Kaplan–Meier curves.

Results: No one with 36 repeats (n = 25) phenoconverted. CP for 38 repeats (n = 120) was 32% (95% confidence interval [CI] 0%–55%) and 51% (CI, 10%–73%) by ages 70 and 75, respectively, and 68% (CI, 46%–81%) and 81% (CI, 58%–92%) by ages 70 and 75 for 39 repeats (n = 253). CP was not estimable at those ages for 37 repeats (n = 33).

Conclusions: Differences by RP-range repeat length did not reach significance with a 3-year median follow-up duration among censored individuals. © 2021 International Parkinson and Movement Disorder Society

Key Words: phenoconversion; CAG repeat length; manifest Huntington's disease; cumulative penetrance; prospective study

Huntington's disease (HD) is a fatal autosomal dominant neurodegenerative disorder caused by a CAG repeat expansion in the huntingtin gene (*HTT*). For those with 40 or more CAG repeats, penetrance is 100%, barring competing risk of death, and several estimates of disease onset have been published.[1,2] Less is known about age of onset for those with repeat lengths of 36 to 39, known as the "reduced penetrance" (RP) range.[3]

Penetrance rates for RP allele carriers have been reported previously. Brinkman et al[4] estimated a 68% risk of neurological or psychological onset by age 70 for those with 39 repeats, but did not evaluate smaller repeat lengths. Langbehn et al[5] estimated risk of motor onset by age 70 to be 10% for 36 repeats, 17% for 37 repeats, 32% for 38 repeats, and 53% for 39 repeats. However, these estimates were extrapolated from a parametric model designed for individuals with 41 to 56 repeats; RP allele carriers were excluded because of potential non-representative ascertainment in this range. Quarrell et al[6] reported an empirical cumulative penetrance (CP) of 67.1% by age 70 for RP alleles; however, results for individual alleles may have been unreliable because of small sample size (n = 176).

Several prospective observational studies emphasized recruitment of pre-manifest individuals at risk for HD, which may alleviate some ascertainment bias. In this paper, we use the most recent data (December 2020) from the largest such study, Enroll-HD, to provide empirical estimates of HD penetrance over time for RP allele carriers and compare to full penetrance (FP) allele carriers.

Methods

Study Population

Data from Enroll-HD were made available by CHDI Foundation Core datasets are collected annually for over 20,000 participants globally. Data are monitored for quality and accuracy using a risk-based monitoring approach. All sites must obtain and maintain local ethical approval. For this analysis we used Periodic Data Set 5, which also included data from REGISTRY, Enroll-HD's European predecessor study. When applicable, we included patients' preceding REGISTRY

*Correspondence to: Dr. E.I. McDonnell, Department of Biostatistics, 722 West 168th Street, New York, NY 10032, USA; E-mail: eim2117@cumc.columbia.edu

Relevant conflicts of interest/financial disclosures: The authors report no conflicts of interest.

Funding agencies: This work was supported in part by the National Center for Advancing Translational Sciences, through grant number TL1TR001875. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This work was also supported in part by the National Institute of Neurological Disorders and Stroke, through grant number R01NS073671.

Received: 14 May 2021; **Revised:** 4 August 2021; **Accepted:** 23 August 2021

Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.28789

(protocol version 3) visits. Individuals were eligible for this analysis if they were HD gene expansion carriers, and if at baseline (Enroll-HD or REGISTRY3, as appropriate) they were determined to be pre-manifest with a diagnostic confidence level (DCL) <4 on the Unified HD Rating Scale (UHDRS).[7]

Statistical Methods

Kaplan–Meier (KM) curves were fit to estimate integer age of phenoconversion to manifest HD diagnosis, defined as the first occurrence of a DCL of 4. Individuals who never reached a DCL of 4 were censored at the age of their last recorded DCL. To account for left-truncation (exclusion of individuals who were manifest at baseline), we used a counting process style KM estimation, only including individuals in the risk set while they were on study (ie, period between baseline age and diagnosis or censoring age). Because of this, the number at risk for KM estimation can increase or decrease over age. There may be periods with only one or even zero people in the risk set, followed by enrollment of new individuals. If an individual develops manifest HD while being the only person in the risk set, the KM curve will estimate 100% CP from that time forward, even if other individuals enter the study at later ages, leading to unreliable estimates. Therefore, we administratively censored data at the earliest age beyond 60 years that the risk set decreased to only one individual. CP at age t, or the probability of phenoconverting by age t, was defined as $F(t) = 1 - S(t)$, where $S(t)$ is the KM survival curve. Log-rank tests were performed at a 5% significance level to compare KM curves by CAG repeat length and sex.

Results

After data cleaning, 3868 HD gene expansion carriers were eligible for analysis, defined as pre-manifest at baseline (DCL <4) with at least one post-baseline DCL on record. Among them, 431 (11%) had CAG repeat lengths in the RP range (n = 25 with 36 repeats, n = 33 with 37 repeats, n = 120 with 38 repeats, and n = 253 with 39 repeats). The remaining 3437 with alleles in the FP range had a median repeat length of 42 (range 40–60). RP allele carriers were older on average at baseline than FP allele carriers (mean ± standard deviation [SD]: 49.2 ± 13.4 vs. 38.6 ± 11.3 years, respectively; t test, P value <0.001). A total of 42% of RP allele carriers and 40% FP allele carriers were male. Baseline DCL was 0 to 1 for 91% of RP allele carriers and 84% of FP allele carriers (χ^2, P value <0.001).

Median follow up duration among censored individuals was 3 years (range 1–9 years). During follow up, 41 RP allele carriers and 820 FP allele carriers developed manifest HD (DCL of 4). After administratively censoring the RP cohort at 79 years and the FP cohort at 76 years, age of manifest HD diagnosis was significantly older in RP versus FP allele carriers (median 66 vs. 38 years; P value <0.001; Fig. 1a). Among RP allele carriers, we administratively censored those with 36, 37, 38, and 39 repeats at ages 60, 61, 77, and 78, respectively. Allele-specific CP probabilities at various ages are presented in Table 1. No one with 36 repeats phenoconverted while on study, even without administrative censoring. For 37, 38, and 39 repeats, CP probabilities (and 95% confidence interval [CI]) by age 60 were 44% (CI, 0%–75%), 32% (CI, 0%–55%) and 43% (CI, 21%–58%), respectively. Among RP allele carriers,

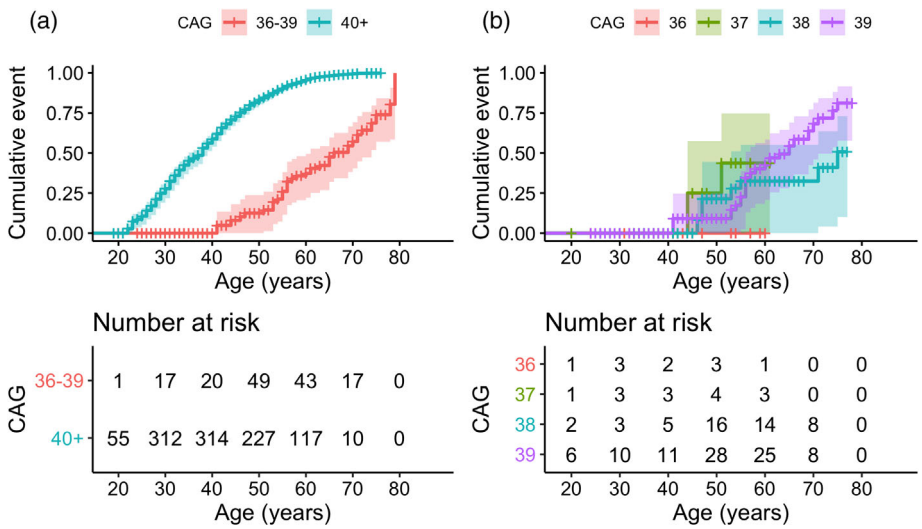

FIG. 1. Cumulative penetrance over age, stratified by CAG repeat length. (**a**) Stratified by CAG repeat length 36–39 versus 40+. Log-rank P value <0.001. (**b**) Individuals with RP alleles only, stratified by individual repeat length. Log-rank P value = 0.1. Comparing only 38 versus 39 repeats: P value = 0.07. Kaplan–Meier curves were administratively censored at the earliest age beyond 60 years that the risk set contained only 1 individual. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 *Estimated cumulative penetrance at various ages, stratified by CAG repeat length*

	Cumulative penetrance (95% CI)					
Age	CAG 36 (n = 21[a])	CAG 37 (n = 27[a])	CAG 38 (n = 116[a])	CAG 39 (n = 250[a])	CAG 36–39 (n = 424[a])	CAG 40 + (n = 3435[a])
50	0.00 (0.00–0.00)	0.25 (0.00–0.57)	0.21 (0.00–0.44)	0.09 (0.00–0.25)	0.13 (0.00–0.24)	0.83 (0.80–0.86)
55	0.00 (0.00–0.00)	0.44 (0.00–0.75)	0.32 (0.00–0.55)	0.23 (0.02–0.39)	0.26 (0.12–0.38)	0.91 (0.89–0.92)
60	0.00 (0.00–0.00)	0.44 (0.00–0.75)	0.32 (0.00–0.55)	0.43 (0.21–0.58)	0.37 (0.22–0.49)	0.96 (0.95–0.97)
65	NA	NA	0.32 (0.00–0.55)	0.54 (0.33–0.69)	0.48 (0.32–0.60)	0.98 (0.98–0.99)
70	NA	NA	0.32 (0.00–0.55)	0.68 (0.46–0.81)	0.57 (0.41–0.69)	0.99 (0.99–1.00)
75	NA	NA	0.51 (0.10–0.73)	0.81 (0.58–0.92)	0.74 (0.57–0.84)	1.00 (1.00–1.00)

[a]Sample size based on administratively censoring data at the first age beyond 60 when only one individual remains in the risk set.
Abbreviations: CI, confidence interval; NA, estimates are not available at the given age because of insufficient data.

we found no significant differences in age of phenoconversion by sex (P value = 0.4) or by individual repeat length (P value = 0.1, Fig. 1b. 38 vs. 39 repeats: P value = 0.07).

Discussion

Enroll-HD contains the largest sample of prospectively collected data on HD penetrance for RP alleles. We leveraged this rich dataset to provide a new set of empirical penetrance estimates for each allele within the RP range.

Comparing our findings to the literature, although we found that no one with 36 repeats reached manifest HD, cases of penetrance for this allele have previously been documented.[6] For individuals with 37, 38, and 39 repeats, allele-specific median ages of onset (65, 75, and 65, respectively) were younger than the predicted medians from the Langbehn et al formula (85, 76, and 69, respectively),[5] with 37 repeats having the largest discrepancy. Our allele-specific empirical CP estimates are generally higher, although some similar, compared to the model-based estimates from Langbehn et al.[5] Our estimates for 38 repeats at ages 70 and 75 (32% and 51%, respectively) are particularly close to those of Langbehn et al[5] (32% and 45%, respectively). Our estimates for 39 repeats at ages 70 and 75 (68% and 81%, respectively) agree with those of Brinkman et al[4] (68% and 79%, respectively), but not Langbehn et al,[5] who reported similar CP probabilities for 5 years later in life (69% and 81% at ages 75 and 80, respectively). Langbehn et al[5] had extrapolated their estimates from parametric models that exclusively used data from FP allele carriers because of concerns about ascertainment bias in the RP range. Our analysis used prospectively-collected data from individuals who were pre-manifest at baseline, so risk of ascertainment bias is low. Furthermore, we mitigated additional bias from left-truncation by only including patients in the risk set while they were on study, rather than from birth. Our findings suggest that some of Langbehn et al's[5] extrapolated predictions of CP may be too low in the RP allele population.

Our findings support claims that age of onset can vary even among individuals with the same CAG repeat length.[1] One proposed explanation for this is simply interlaboratory variation in repeat length determination.[8] However, Enroll-HD used a central biorepository facility to conduct all genotyping. There is increasing evidence that the length of the uninterrupted CAG repeat length is what correlates with age at onset. A loss of interruption (LOI) variant has been shown to associate with much earlier age of onset, particularly in RP allele carriers. In one study, among RP allele carriers, the LOI was found in 33.3% of symptomatic individuals, versus 5.1% of asymptomatic individuals. Prevalence of the LOI was 100% among symptomatic individuals with 36–37 repeats. The LOI was associated with a striking 29.1-year earlier age of onset than predicted for RP allele carriers, and a 12.8-year earlier age of onset than predicted for FP allele carriers.[9] In a second study, the LOI, which was found in 1.02% of symptomatic HD gene expansion carriers and 32.2% of symptomatic RP allele carriers, was associated with a 9.5-year earlier HD onset than expected after correcting for possible underestimation of CAG repeats because of the LOI.[10] The likely mechanism is increased somatic instability, which has been linked to earlier age of onset in FP allele carriers.[11] LOI data are not yet available in Enroll-HD, but it is possible that all two phenoconverters with 37 repeats in our KM analysis and some phenoconverters with 38 to 39 repeats may have the LOI variant. This would explain the earlier median age of onset than predicted for 37 repeats. Future work is needed to estimate the effect of LOI on penetrance in this population.

This analysis is clinically important for genetic counseling of individuals from known HD families with

repeat lengths in the RP range who want more definitive information of their risk of phenoconversion.[11] It is also important for the general population. One study estimated that 1 in 400 people in the general population have a CAG repeat expansion ≥36, most of which are in the RP range.[12]

This analysis has several limitations. First, RP and FP populations differed in age at study entry. It may be more difficult in the older RP sample to discern HD symptoms from signs of normal aging during UHDRS evaluation. Second, although Enroll-HD specifically recruited pre-manifest HD gene expansion carriers for prospective observation, ascertainment bias may still be present if the study populations are only representative of those seeking medical care.[3] Future work is needed to include community controls in an analysis of HD penetrance. Third, although visits were scheduled annually, some individuals had gaps of 2 to 9 years between visits, posing risk of interval censoring for some phenoconverters. Fourth, we assumed all individuals were independent, although some correlation between family members may exist. Last, future work is needed to incorporate death as a competing risk. ■

Acknowledgment: Biosamples and data used in this work were generously provided by the participants in the Enroll-HD study and made available by CHDI Foundation. Enroll-HD is a clinical research platform and longitudinal observational study for Huntington's disease families intended to accelerate progress toward therapeutics; it is sponsored by CHDI Foundation, a nonprofit biomedical research organization exclusively dedicated to collaboratively developing therapeutics for HD. Enroll-HD would not be possible without the vital contribution of the research participants and their families. We acknowledge the individuals who contributed to the collection of the Enroll-HD data (listed at https://www.enroll-hd.org/acknowledgments/).

Data Availability Statement

The fifth Enroll-HD Periodic Data Set (PDS5) is made available by CHDI Foundation, Inc. Verified researchers may request access at the following link: https://enroll-hd.org/for-researchers/access-data/.

References

1. Langbehn DR, Hayden M, Paulsen JS. The PREDICT-HD investigators of the Huntington study group. CAG-repeat length and the age of onset in Huntington disease (HD): a review and validation study of statistical approaches. Am J Med Genet B Neuropsychiatr Genet 2010;153B(2):397–408.
2. Chen T, Wang Y, Ma Y, Marder K, Langbehn DR. Predicting disease onset from mutation status using proband and relative data with applications to Huntington's disease. J Probab Stat 2012; 2012:375935.
3. The American College of Medical Genetics/American Society of Human Genetics Huntington Disease Genetic Testing Working Group. ACMG/ASHG statement: laboratory guidelines for Huntington disease genetic testing. Am J Hum Genet 1998;62(5): 1243–1247.
4. Brinkman RR, Mezei MM, Theilmann J, Almqvist E, Hayden MR. The likelihood of being affected with Huntington disease by a particular age, for a specific CAG size. Am J Hum Genet 1997;60(5): 1202–1210.
5. Langbehn DR, Brinkman RR, Falush D, Paulsen JS, Hayden MR. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. Clin Genet 2004;65(4): 267–277.
6. Quarrell OWJ, Rigby AS, Barron L, Crow Y, Dalton A, Dennis N, et al. Reduced penetrance alleles for Huntington's disease: a multi-Centre direct observational study. J Med Genet 2007;44(3):e68.
7. Huntington Study Group. Unified Huntington's disease rating scale: reliability and consistency. Mov Disord 1996;11(2):136–142.
8. Rubinsztein DC, Leggo J, Coles R, Almqvist E, Biancalana V, Cassiman J-J, et al. Phenotypic characterization of individuals with 30-40 CAG repeats in the Huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36-39 repeats. Am J Hum Genet 1996;59(1):16–22.
9. Wright GEB, Collins JA, Kay C, McDonald C, Dolzhenko E, Xia Q, et al. Length of uninterrupted CAG, independent of polyglutamine size, results in increased somatic instability, hastening onset of Huntington disease. Am J Hum Genet 2019;104(6):1116–1126.
10. Findlay Black H, Wright GEB, Collins JA, Caron N, Kay C, Xia Q, et al. Frequency of the loss of CAA interruption in the HTT CAG tract and implications for Huntington disease in the reduced penetrance range. Genet Med 2020;22(12):2108–2113.
11. Migliore S, Jankovic J, Squitieri F. Genetic counseling in Huntington's disease: potential new challenges on horizon? Frontiers in Neurology 2019;10:1–6.
12. Kay C, Collins JA, Miedzybrodzka Z, Madore SJ, Gordon ES, Gerry N, et al. Huntington disease reduced penetrance alleles occur at high frequency in the general population. Neurology 2016;87(3): 282–288.

SGML and CITI Use Only
DO NOT PRINT

Author Roles

(1) Research project: A. Conception, B. Organization, C. Execution; (2) Statistical Analysis: A. Design, B. Execution, C. Review and Critique; (3) Manuscript: A. Writing of the First Draft, B. Review and Critique.

E.I.M.: 2A, 2B, 3A
Y.W.: 2A, 2C, 3B
J.G.: 1A, 3B
K.M.: 1A, 2A, 3B

Financial Disclosures

Financial disclosures for all authors (for the preceding 12 months)

This work was supported in part by the National Center for Advancing Translational Sciences, National Institutes of Health (NIH), through grant number TL1TR001875. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

E.I.M. has received support from the National Institute of Neurological Disorders and Stroke, through grant numbers R01NS097423 and R01NS073671. Y.W. is supported by the National Institute of Neurological Disorders and Stroke, grant number R01NS073671. K.M. received support from the CHDI Foundation, Huntington's Disease Society of America, Huntington Study Group, site investigator for Genentech, Triplet Therapeutics, and Prilenia Therapeutics. She receives research support from the NIH (NS100600, UL1TR001873, U24NS107168), Parkinson Disease Foundation, The Michael J. Fox Foundation, and Lewy Body Disease Association.