



OPEN

Sex contribution to average age at onset of Huntington's disease depends on the number of (CAG)_n repeats

Anna Stanisławska-Sachadyn^{1,2,3✉}, Michał Krzemiński⁴, Daniel Zielonka⁵, Magdalena Krygier⁶, Ewa Ziętkiewicz⁷, Jarosław Sławek^{8,9}, Janusz Limon¹⁰ & REGISTRY investigators of the European Huntington's Disease Network (EHDN)*

Huntington's disease (HD) is a hereditary neurodegenerative disorder caused by the extension of the CAG repeats in exon 1 of the *HTT* gene and is transmitted in a dominant manner. The present study aimed to assess whether patients' sex, in the context of mutated and normal allele length, contributes to age on onset (AO) of HD. The study population comprised a large cohort of 3723 HD patients from the European Huntington's Disease Network's REGISTRY database collected at 160 sites across 17 European countries and in one location outside Europe. The data were analyzed using regression models and factorial analysis of variance (ANOVA) considering both mutated allele length and sex as predictors of patients' AO. AO, as described by the rater's estimate, was found to be later in affected women than in men across the whole population. This difference was most pronounced in a subgroup of 1273 patients with relatively short variants of the mutated allele (40–45 CAG repeats) and normal alleles in a higher half of length distribution—namely, more than 17 CAG repeats; however, it was also observed in each group. Our results presented in this observational study point to sex-related differences in AO, most pronounced in the presence of the short mutated and long normal allele, which may add to understanding the dynamics of AO in Huntington's Disease.

Trial registration: *ClinicalTrials.gov* identifier NCT01590589.

Huntington's disease (HD) is an autosomal dominant, progressive, neurodegenerative disorder characterized by motor symptoms, cognitive impairment, and psychiatric disturbances. The prevalence of HD is 6.37 and 8.87 per 100,000 people in Europe and North America, respectively, while it is much less prevalent in Africa and Asia¹. It is caused by an expansion of the trinucleotide CAG tandem repeat (> 35 CAGs) located in exon 1 of the *HTT* gene encoding the huntingtin protein. Huntingtin's cellular functions are not fully understood, although mutant huntingtin is known to be digested into fragments, which build toxic aggregates that disrupt transcription processes and cytoplasmic transport². This leads to mitochondrial dysfunction, altered reactive oxygen species defense, and finally, apoptosis resulting in neuronal dysfunction.

The onset of the disease usually occurs in midlife. Age at onset (AO) is strongly associated with the length of the (CAG)_n expansion in the mutated allele, such that the longer the expanded repeat is, the earlier the onset

¹Department of Biotechnology and Microbiology, Gdańsk University of Technology, 80-233 Gdańsk, Poland. ²Department of Biology and Medical Genetics, Medical University of Gdańsk, 80-211 Gdańsk, Poland. ³BioTechMed Center, Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland. ⁴Institute of Applied Mathematics, Gdańsk University of Technology, 80-233 Gdańsk, Poland. ⁵Department of Public Health, Poznań University of Medical Sciences, 60-812 Poznań, Poland. ⁶Department of Developmental Neurology, Medical University of Gdansk, 80-952 Gdańsk, Poland. ⁷Institute of Human Genetics, Polish Academy of Sciences, 60-479 Poznań, Poland. ⁸Department of Neurology, St. Adalbert Hospital, Copernicus PL, 80-462, Gdańsk, Poland. ⁹Department of Neurological and Psychiatric Nursing, Faculty of Health Sciences, Medical University of Gdańsk, 80-211 Gdańsk, Poland. ¹⁰Department of Medical Ethics, Medical University of Gdańsk, 80-211 Gdańsk, Poland. *A list of authors and their affiliations appears at the end of the paper. ✉email: anna.stanislawski@pg.edu.pl

of clinical symptoms will manifest. Mutated alleles that carry at least 40 repeats are fully penetrant³. Expanded alleles ranging from 40 to 45 repeats are observed most often.

A considerable variance in AO has been observed in individuals who carry the (CAG)_n expansion at comparable lengths, prompting the search for modifying factors that may account for the varying AO. The length of CAG repeats is responsible for approximately 50–70% of the variability of AO in HD².

The mutated *HTT* is highly unstable, and both extensions and contractions of the mutated allele in inter-generational transmissions are observed in HD⁴. Interestingly, the sex of a mouse embryo has an impact on the likelihood of CAG repeat contractions or expansions, which has led to suggestions that the change in CAG repeat number may also be a post-zygotic event^{5,6}. In humans, it has been observed that the mutated allele may contract when inherited from mother^{7,8}. Notably, 42% of maternally transmitted alleles contracted when passed to daughters and only 27% when passed to sons⁹. The above observation led to a suggestion that both parental and offspring sex plays a role in HD pathology. Interestingly, data regarding the impact of sex on AO of HD are relatively scarce^{10–13}, and it is clear that the involvement of the patient's sex in the AO of HD should be further evaluated in the context of the size of the mutated allele.

In recent years it has been noticed that the expansion of CAG triplets occurs not only in germinal but also in somatic cells, including the nervous system—specifically, in human striatal cells—and is an early event in the disease course¹⁴. CAG trinucleotide instability has been observed in non-dividing cells^{15,16}, it has been demonstrated that the expansions of a mutated allele in the cortex¹⁷ and in the post-mitotic striatum¹⁶ determine the AO of HD.

The mechanism by which expansions arise may be a consequence of the physiological processes present in somatic cells¹⁷. It has been recently established that more genetic factors are involved in the mechanism by which the expansion in non-dividing cells arises. Somatic instability is dependent on the presence of mismatch repair proteins—namely Msh2¹⁸, Msh3 and Mlh1¹⁹, Msh2-Msh3 form MutS-beta heterodimer recognizing and binding CAG loops. In the genome-wide association study (GWAS), other DNA repair system genes—the *FAN1* and the *RRM2B*—have been identified as genetic factors that underlie variation in AO²⁰. The *FAN1* gene encodes a nuclease involved in inter-strand DNA crosslink repair that is necessary for controlling CAG repeat expansion²¹. Another GWAS study confirmed that the *MLH1* gene is a factor modifying AO in HD²². It has been further hypothesized that the mismatch repair system (MMR) may introduce expansion while correcting for misalignment of DNA strands during its reannealing, either in the course of DNA replication in dividing cells or RNA transcription in non-dividing cells, in a process initialized by binding of PCNA and MutS-beta²³.

Apart from the mutated allele, the role of a normal allele in HD pathogenesis has been also considered. The negative correlation between AO and the normal CAG repeat number has been reported to be sex-specific and stronger for normal maternal allele transmissions²⁴. Irrespective of the patient's sex a negative correlation between AO and the normal paternal allele size has also been reported²⁵. Interestingly, results of analyses of 337 inter-generational transmissions indicate that, in patients who inherit the expanded allele from their mothers, the increased frequency of mutated allele contractions has been associated with longer normal alleles²⁶, pointing at differences in the interplay between mutated and normal alleles between sexes in trans-generational length changes. Next, larger normal alleles in combination with shorter mutated alleles were associated with earlier AO and vice versa—in combination with longer mutated alleles—with later AO, in a group of 921 subjects²⁷. Similarly, the longest normal alleles were linked to later AO when combined with long expanded alleles²⁸. Nevertheless, no impact of the normal allele length on motor onset along the mutated allele length was found in multiple linear regression analyses in a large population of 4067 subjects²⁹.

The present understanding of the impact of a variety of factors on AO of HD remains not fully solved. The present study aimed to test the extent to which the sex of affected individuals can impact AO of HD in the context of the length of CAG repeats within large and normal alleles. The analyses were conducted in a large-scale, multi-national cohort of European ancestry. This observational study supported by statistical analyses adds to understanding the pathogenesis of HD in a way that could not be achieved in animal or cellular models.

Methods

Study population

Analyses were performed on data extracted from the REGISTRY database provided by the European Huntington's Disease Network (EHDN). The data were obtained as part of the EDHN's data mining project 0636. REGISTRY data were collected at 160 sites across in 18 countries (17 European and one country outside Europe) from June 2004 and were assessed in November 2017. Given that the data required to perform the analyses were not available for all the HD patients included in the REGISTRY database, the selection of the study population was necessary. The criteria for including subjects were (i) the number of (CAG)_n repeats in the expanded allele equal to or higher than 36 (10,363 subjects) and (ii) the availability of the clinician's best estimate of AO—referred to as the rater's estimate that reflects most probable onset age based on experienced professional interview with patient and family members or onset established examining person observed from premanifest HD stage, the number of (CAG)_n repeats in the normal allele, if available, lower than 36 (3723 subjects). The rater's estimate of AO was calculated based on the "srxrater," which is coded as a date in the REGISTRY database. The EHDN investigators' AO estimation was used in analyzes. The data on the (CAG)_n allele length were from the EHDN database or, if not available, from the local laboratory. The number of repeated CAG units in the expanded allele ranged from 36 to 90. AO was analyzed in the context of patients' sex in a group of 3723 patients (see Table 1 for further details).

Ethical approval for REGISTRY was obtained in each participating country. All participants gave written informed consent: https://www.enroll-hd.org/enrollhd_documents/2016-10-R1/registry-protocol-3.0.pdf. The REGISTRY protocol was approved by the EHDN Scientific and Bioethics Advisory Committee.

Statistical analyses

Multiple regression models to estimate the variance in patients' AO were created. The regression coefficients from the analyzed models were used to assess the proportion of variance in patients' AO (the outcome measure) explained by mutated allele length and the other predictor variable: the sex of HD patients or a type of first symptoms. Regression coefficients were calculated while the expanded allele sizes were median-centered for 43 CAG repeats. The interaction terms between the mutated allele length and patients' sex were not significant in any of the models (Table 2, models A–F); therefore, models without interaction terms were chosen. In multiple linear regression models designed to determine whether the impact of a first symptom type differed between sexes, the significance of the interaction coefficients term between the patient's sex and type of first symptoms were of interest.

In simple linear regression analyses, the coefficient of determination (R^2) was used to assess the proportion of variance in AO explained by the predictor variable.

A two-way ANOVA using a generalized linear model with the LSMEAN statement and Tukey post-hoc test was performed. The association between AO and the number of (CAG)_n repeats in the mutated allele with the impact of patients' sex was assessed using factorial analyses with a two between-subjects factor.

In both the factorial and regression analyses, AO, which was a dependent variable, was natural log-transformed³⁰. The normality of the distributions was assessed using the Kolmogorov–Smirnov test.

In the ranked analyses, the significance of differences between the groups was determined using the Wilcoxon rank-sum test, corrected for multiple comparisons using the Bonferroni method when necessary. Chi-square statistics were used to test the difference in distribution of the categorical variables (e.g., frequency of first symptom types between sexes). Correlations between the number of CAG repeats within either normal or mutated allele, and the AO values were calculated using Spearman statistics.

The level of significance was set at 0.05. The statistical analyses were carried out using SAS versions 9.3 and 9.4 (NC, USA) and the R platform.

Population stratifications

Several analyses were performed in subgroups defined according to the ranges of the expanded allele size: ≤ 39 (CAG)_n, 40–45 (CAG)_n, 46–50 (CAG)_n, and > 50 (CAG)_n. Subjects in group 1 had alleles of 39 CAG repeats or fewer, which are not fully penetrant³. Carriers of penetrant and mid-size mutation of 40–50 CAGs had an average age at onset HD. They are the majority of HD patients whose disease is characterized by a large diversity in AO; thus, it was further divided into group 2, including subjects who had expanded alleles of 40–45 CAGs, and group 3 who had expanded alleles of 46–49 CAGs. Subjects in group 4 carried mutation longer than 50 CAGs and had juvenile form of HD.

Expanded alleles described 27.85% of the variance in AO among 2617 carriers of 40–45 CAGs (70.25% of those for whom sex, AO, and mutated allele length were available), 10.68% among 757 carriers of 46–50 CAGs, and 60.04% among 255 carriers of alleles longer than 50 CAGs, who established an early-onset subgroup. Thus, the coefficient of determination (R^2) in regression models, which is used to assess the proportion of variance in AO explained by the expanded allele size in the whole cohort is highly influenced by the characteristics of early-onset subjects, for whom genetic factor plays the strongest role in HD pathogenesis. Thus, the precision in determining the influence of an additional factor on AO, specifically of the patient's sex, increases in analyses conducted among subgroups of mutated allele sizes, which certainly have a dominant impact.

Moreover, in several analyses, the population has been stratified into two groups regarding the size of a normal allele, that is, below/equal to and above a median value of 17 CAGs. Subjects in a group of lower normal allele length distribution carried 8–17 CAG repeats, while subjects in a group of higher normal allele length distribution carried 18–35 CAG repeats.

Results

Data distribution and basic statistics

A clinician's best estimate of AO (i.e., rater's estimated AO) and expanded allele length were available for 3723 subjects: men accounted for 47.95% and women for 52.05% (Table 1). Data pertaining to normal alleles were available for 3643 subjects. The frequencies of the first symptom types, as listed in rater's estimate, were comparable between women and men ($P = 0.2294$, Table 1), although a slightly higher proportion of men first observed motor (52.82%) and cognitive symptoms (7.90%) compared to women (51.48% and 6.71%, respectively), while more women first observed mixed symptoms (19.76%) than men did (17.95%). These differences between sexes were non-significant, whether assessed using chi-square statistics (Table 1) or multiple regression analyses, which showed that the interaction term between the patient's sex and the first symptom types did not contribute significantly to explaining the proportion of variance in AO (data not presented).

A simple group comparison revealed no statistically significant difference in AO between men and women (Table 1). The median AO was 44 years for men and 45 for women ($P = 0.0793$), while the median length of the expanded allele was 43 for both men and women ($P = 0.4376$).

As expected, the mutation size correlated highly with AO both in women and men ($r = -0.75863$, $r = -0.76530$, respectively; both $P < 0.0001$). In the study population analyzed as a whole, the size of a mutated allele described 60.65% of the variance in AO—61.30% in women and 60.18% in men.

Association between patients' sex and age at onset in the context of mutated allele length

In the whole study population, in regression models describing the proportion of variance in AO and involving the mutated allele length and patient sex as two predictor variables, sex was significantly involved in the AO variability ($P = 0.0012$; Table 2A). After the cohort was divided with respect to mutation size, this association stayed significant among those subjects who had 40–45 CAG repeats ($P = 0.0006$; Table 3B) but not in those who had ≤ 39 (CAG)_n, 46–50 (CAG)_n and > 50 (CAG)_n (data not presented).

	All	Men	Women	P
				Males vs females
¹ Study group	N = 3723	N = 1785 (47.95%)	N = 1938 (52.05%)	
Age at onset (rater's estimate)				
Median (q1–q3)		44 (35–52)	45 (36–53)	0.0793 [#]
Mean, SD		43.96, 12.29	44.61, 12.30	
First symptoms				
1 (motor)		N = 936 (52.82%)	N = 990 (51.48%)	0.2294 [*]
2 (cognitive)		N = 140 (7.90%)	N = 129 (6.71%)	
3 (psychiatric)		N = 358 (20.20%)	N = 402 (20.90%)	
4 (oculomotor)		–	N = 3 (0.16%)	
5 (other)		N = 20 (1.13%)	N = 19 (0.99%)	
6 (mixed)		N = 319 (17.95%)	N = 380 (19.76%)	
Expanded allele length				
Median (q1–q3)		43 (42–46)	43 (42–46)	0.4376 [#]
Mean, SD		44.31, 4.26	44.41, 4.41	
² Subgroup with data on normal allele length	N = 3643	N = 1749	N = 1894	
Normal allele length				
Median (q1–q3)		18 (17–20)	17 (17–19)	0.1431 [#]
Mean, SD		18.49, 3.19	18.41, 3.28	

Table 1. Study population characteristics, with regards to HD patients' sex; for the whole study group and among those for whom both alleles within the *HTT* gene were known. ¹Data restricted to subjects whose expanded allele \geq CAG₃₆, normal allele $<$ CAG₃₆ (if available), for whom sex, AO, number of (CAG)_n was available. Homozygous subjects (genotype: 42/37, 44/36, 44/39, 45/45, 49/39, 54/37, 54/37, 65/36) were excluded. ²Data restricted to subjects for whom numbers of (CAG)_n within both alleles were available. [#]Wilcoxon Rank Sum Test. *Chi-square.

Intercept (P-value) ¹	¹ Mutated allele (P-value) ¹	² Sex of HD patient (P-value)	Interaction between mutation and patient's sex (P-value)	R ² (P-value) ¹	N
(A) Whole population					
3.81443	– 0.05896	0.01852 (0.0076)	0.00222 (0.1464)	0.60781	3723
3.81288	– 0.05776	0.02152 (0.0012)	–	0.60759	
(B) Patients with mutated allele of 40–45 CAG					
3.81690	– 0.07993	0.02734 (0.0003)	0.00510 (0.2935)	0.28211	2617
3.81802	– 0.07732	0.02526 (0.0006)	–	0.28180	
(C) Patients with normal allele $>$ 17 CAG					
3.80836	– 0.05743	0.02937 (0.0045)	0.00263 (0.2352)	0.59334	1802
3.80635	– 0.05589	0.03312 (0.0007)	–	0.59302	
(D) Patients with normal allele \leq 17 CAG					
3.81982	– 0.06022	0.01041 (0.2738)	– 0.00014 (0.9468)	0.62755	1841
3.81991	– 0.06028	0.01023 (0.2615)	–	0.62755	
(E) Patients with normal allele $>$ 17 CAG, with mutated allele of 40–45 CAG					
3.80952	– 0.07574	0.04584 ($<$ 0.0001)	0.00653 (0.3746)	0.24131	1273
3.81080	– 0.07252	0.04340 ($<$ 0.0001)	–	0.24084	
(F) Patients with normal allele \leq 17 CAG, with mutated allele of 40–45 CAG					
3.82523	– 0.08362	0.01027 (0.3209)	0.00180 (0.7843)	0.33226	1285
3.82564	– 0.08269	0.00952 (0.3398)	–	0.33222	

Table 2. Multiple linear regression models describing the proportion of variance in AO explained by mutated allele and sex of HD patient, either with or without interaction between mutation and patient's sex. Independent variables were median-centered, thus an intercept equals $\ln(\text{AO})$ while mutated allele equals 43. ¹ $P < 0.0001$. ²Males were reference sex.

Similarly, in two-way ANOVA LSMEAN analyzes, patients' sex had a significant impact on AO variability across the entire study population ($P=0.0004$; Table 3A, Fig. 1 A). In those analyzes, unlike in the regression models, the mean AO is assessed by ANOVA between women and men for each mutated allele of the same size. Since no interaction between the length of a mutated allele and patients' sex was found, analysis of type II was

Impact of expanded allele and sex on age at onset in HD						
Factor	DF	Type II SS	Mean square	F value	P-value	N
(A) Whole population						
Mutated allele	41	244.2525553	5.9573794	155.63	<0.0001	3723
Sex	1	0.4797997	0.4797997	12.53	0.0004	
Mutated allele*sex interaction	32	1.3539487	0.0423109	1.11	0.3131	
(B) Mutated allele of $(CAG)_n = 40-45$						
Mutated allele	5	35.81407630	7.16281526	204.28	<0.0001	2617
Sex	1	0.42192903	0.42192903	12.03	0.0005	
Mutated allele*sex interaction	5	0.22615282	0.04523056	1.29	0.2653	
(C) Normal allele of $(CAG)_n > 17$						
Mutated allele	36	119.6907953	3.3247443	81.97	<0.0001	1802
Sex	1	0.5230997	0.5230997	12.90	0.0003	
Mutated allele*sex interaction	25	1.2288383	0.0491535	1.21	0.2157	
(D) Normal allele of $(CAG)_n \leq 17$						
Mutated allele	36	122.7888247	3.4108007	95.85	<0.0001	1841
Sex	1	0.0729239	0.0729239	2.05	0.1525	
Mutated allele*sex interaction	24	1.3786835	0.0574451	1.61	0.0303	
(E) Normal allele of $(CAG)_n > 17$ and mutated allele of $(CAG)_n = 40-45$						
Mutated allele	5	14.95028510	2.99005702	77.82	<0.0001	1273
Sex	1	0.59534683	0.59534683	15.50	<0.0001	
Mutated allele*sex interaction	5	0.05386317	0.01077263	0.28	0.9240	
(F) Normal allele of $(CAG)_n \leq 17$ and mutated allele of $(CAG)_n = 40-45$						
Mutated allele	5	20.42851681	4.08570336	129.15	<0.0001	1285
Sex	1	0.03410713	0.03410713	1.08	0.2993	
Mutated allele*sex interaction	5	0.27334713	0.05466943	1.73	0.1252	

Table 3. Impact of the sex and mutated allele length on the AO in patients, analyzed by two-way ANOVA. The analyses were carried (A) across the entire study group, and in the following subgroups: (B) in subjects with the mutated allele of $(CAG)_n = 40-45$ (C) in subjects with the normal allele of $(CAG)_n > 17$ (D) in subjects with the normal allele of $(CAG)_n \leq 17$ (E) in subjects with the mutated allele of $(CAG)_n = 40-45$ and normal allele of $(CAG)_n > 17$ (F) in subjects with the mutated allele of $(CAG)_n = 40-45$ and normal allele of $(CAG)_n \leq 17$.

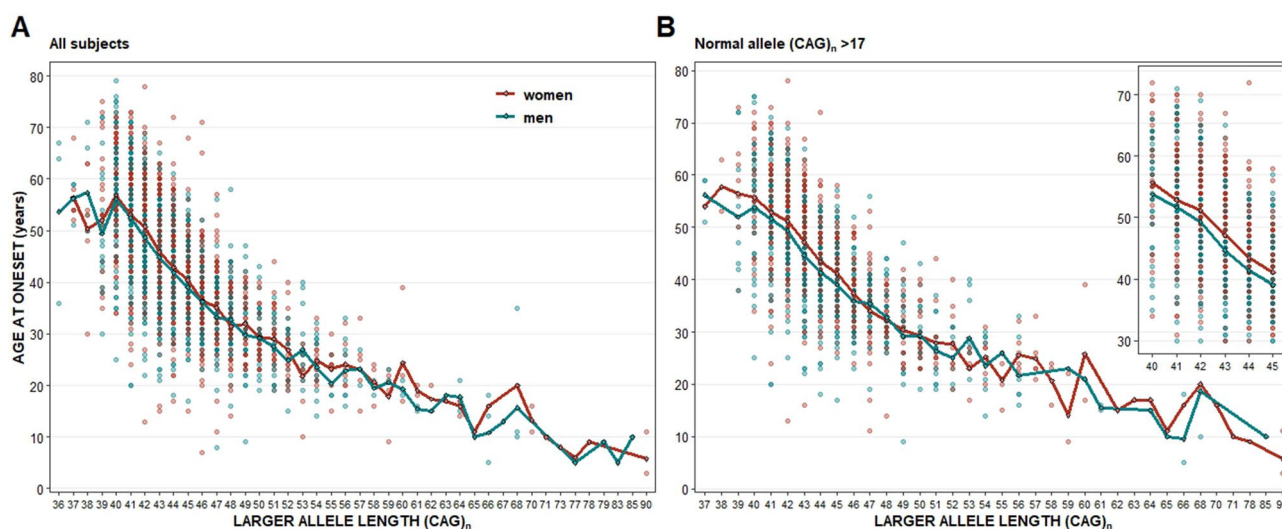


Figure 1. Age at onset of male and female patients in relation to the number of CAG repeats within the mutated allele (A) across the entire study group (B) among subjects who carry more than 17 CAG repeats within the normal allele (higher half of the length distribution), inset: analyses for subjects with mutated allele $(CAG)_n = 40-45$. Two-way ANOVA and regression analyses were performed with natural log-transformed AO.

chosen³¹. In analyses performed in subgroups defined by the range of mutated allele length, patient sex had a significant impact on AO variance ($P=0.0005$; Table 3B) among those subjects who had 40–45 CAG repeats.

The subgroup of HD patients who had expanded allele of 40–45 CAG repeats was characterized by high variance in AO, ranging from the age of 5 to 83; the median AO was 48 years (mean \pm SD; 48.68 ± 9.99). Analyses in ranges to evaluate differences in AO between women and men in the context of mutated allele length are presented in Supplementary Table S1.

Association between patients' sex and age at onset in the context of the length of both mutated and normal alleles

To evaluate whether the association between AO and patient sex might be impacted by the number of repeats within a normal allele, the study population has been stratified by the normal allele length ranges into the lower or higher half of distribution.

In the multiple regression models, patient sex was a significant factor among those who had a normal allele in the higher half of length distribution, that is, 18–35 CAG repeats ($P=0.0007$; Table 2C) but not among those who had normal allele equal to or shorter than 17 CAG repeats ($P=0.2615$; Table 2D).

When the population was stratified according to both mutated and normal allele length ranges, the results remained significant among those who had a mutated allele of 40–45 CAGs and a normal allele in the higher half of the length distribution ($P<0.0001$; Table 2E). Interestingly, in this group, sex accounted for 0.80% of variation in AO ($P=0.0014$), as assessed by simple regression analyses (data not presented). Given the complex nature of AO distribution in HD, this association appears to be quite meaningful for a binary factor such as patients' sex.

Analogously, in factorial two-way ANOVA LSMEAN analyses, the difference in AO between affected women and men was significant only among those subjects whose normal allele length was in the higher half of the length distribution (Table 2C, Fig. 1B). This association was particularly pronounced in patients who also had 40–45 CAG repeats in the mutated allele and normal allele in a higher half of allele distribution ($P<0.0001$; Table 3E, Fig. 1B, intercept).

Overall, the mean AO in women and men in the subgroup of 40–45 CAG repeats in the mutated allele and 18–35 CAG repeats in the normal allele was 49.40 ± 9.64 and 47.70 ± 10.03 years, respectively ($P=0.0071$). The difference in mean AO between females and males varied from 0.55 years for those with the mutated allele of 40 CAG repeats—to 2.23 years for those with the mutated allele of 44 CAG repeats. Analyses in ranges to evaluate differences in AO between women and men in the context of both mutated and normal allele length are presented in Supplementary Table S2.

Discussion

Expansion of CAG repeats is a well-documented HD-determining factor that was first identified during the 1990s. AO is negatively correlated with increasing number of repeats in the mutated allele, with the longest alleles (> 50 CAG repeats) determining a juvenile-onset disease. However, in patients with shorter mutated alleles a high variability in AO is observed. This has prompted a search for factors that may contribute to the variance in AO of HD.

We observed the impact of patients' sex on AO of HD, in analyses that exclusively included subjects for whom the rater's estimation of AO was known. Analysis of the entire study population revealed that patients' sex was a significant factor in the variation in AO. However, among those who carried 40–45 CAG repeats, females' AO was distinctly later than that of males. This finding comes as a surprise because earlier such observation has never been confirmed. The first estimation of sex involvement was made before the measurement of mutated allele size became a clinical practice¹³. No difference in AO between women and men among 2068 subjects¹⁰, among 151 subjects³² or in residual AO of diagnostic motor signs, in the study including 4793 subjects¹¹, has been identified through group comparison analyses. Moreover, no difference in age of clinical HD diagnosis between sexes, among 2145 HD patients, was found when the analysis was controlled for the specially invented score, calculation of which involved the individual's current age and repeat length¹². It should be acknowledged that the group examined in the present study was of a relatively large size of 3723 subjects. A further advantage of our work was the use of the clinician's (the rater's) best estimate of AO and the type of statistical analyses that we applied. Although residual AO calculation involves CAG repeat length¹¹, two-way ANOVA compares the AO of women and men for each mutated allele while multiple regression analyses include mutated allele as a linear variable. Our initial observation made in the analyses of the whole study population was followed by the selection of a group among whom the difference in AO reaches higher significance.

Given that the poly-Q mutation in the *HTT* is highly dynamic in brain tissue, leading to its mosaicism^{14,16,17,33–35}, our results may suggest that either the enlargement of expansion size or the prevalence of enlargements may be higher in terminally differentiated male neurons; alternatively, the increase in contraction size or its prevalence might be higher in terminally differentiated female neurons. If so, such somatic sex-related differences in repeat instabilities may underlie our observation. Since the inter-generational and somatic changes might follow a similar direction, the data regarding inter-generational changes in expansion length may add to explaining the possible mechanism behind our observation. When parents' and offspring's sexes were considered in inter-generational analyses, mother-to-daughter contractions were reported^{7,9}. Moreover, in mice sired by the same fathers, sex-related differences in expansion stability have been reported since expansions of the mutated allele were more frequent in males, while contractions were more frequent in females⁵.

Also, several genetic variants have been found to affect AO of HD in a sex-related manner. Recently, variants within the *MSH3/DHFR* locus—a proven source of the dynamic mutation instability in finally differentiated brain tissue¹⁹—have been found to be stronger modifiers of the diagnostic motor signs in women when analyses were set as sex-specific¹¹. Moreover, X-chromosome-wide association study (XWAS) found a variant close to the moesin gene potentially modifying AO³⁶. Differences in AO of HD between genders have previously been

reported for carriers of different genotypes within the *APOE* gene²⁴. We may speculate that those or other variants might be associated with sex-specific differences in neuronal mosaicism and thus disease AO.

A neuroprotective effect of 17 β -estradiol has been reported in rats carrying mutations of 51 CAG repeats³⁷. It has been demonstrated that in neuroblastoma cells estrogen enhances the expression of huntingtin and neuroglobin, which are then complexed and bind to mitochondria in response to H₂O₂ stress to prevent apoptosis; these processes are abandoned in the case of mutated huntingtin³⁸. Moreover, higher mtDNA levels have been observed in leukocytes from women with HD compared to men with HD³⁹, which may be associated with a protective effect against oxidative stress in females.

Our findings suggest that different factors may be associated with AO between subjects who carry 40–45 CAG repeats and have average AO and those who carry longer mutated alleles and thus have earlier AO. It is well-established that juvenile HD and average-onset HD differ at the molecular level: neurons with intranuclear inclusions composed of huntingtin are observed more frequently in juvenile patients, while extracellular structures with the morphology of dystrophic neuritis dominate in the central nervous system of adult-onset patients⁴⁰.

Further, for subjects who carried a normal allele in the higher half of its length distribution (i.e., more than 17 CAG repeats), the difference in AO between sexes was even more pronounced. The impact of the normal allele length on HD pathogenesis has been considered previously^{24,25,27,28,41}. Interestingly, inter-generational contractions were more frequent when mothers carried long normal alleles²⁶. Normal allele length has been demonstrated to impact AO in a manner dependent on the number of repeats in the mutated allele²⁷, with long normal alleles reducing AO in the presence of the short mutated alleles. These data lead us to repeat the speculation that expansion occurs more frequently in male somatic cells of brain tissue while contractions occur more frequently in female cells—specifically, in the presence of the short mutated and long normal allele, leading to us observing the difference in AO between women and men in the presence of this constitutive blood genotype. This hypothesis, however, requires further investigation.

We conducted our analyses in a large study group comprising 3723 subjects from the EHDN REGISTRY database, which is a multi-center, multi-national, prospective, observational study of HD. The diversity of study population combined with the standardized data from patients' examinations is the present study's strength. However, the estimation of AO constitutes a weakness of our study. The rater's confidence level was high for 69.91% of AO estimations in our study group. However, AO was estimated retrospectively, the estimations might have been based on family reports, sometimes several years after the onset, which could lead to a bias in AO. Moreover, the data on the CAG number within both the expanded and normal alleles have been determined in several centres, including the EHDN and local laboratories, which might contribute to a bias in our results.

In conclusion, we have presented analyses indicating that AO depends on the patient's gender with regard to the sizes of both the mutated and the normal alleles. For patients who carry 40–45 CAG repeats, AO occurred later in females than in males. This association was stronger when the normal allele was in the upper range of the size distribution (that is, it was longer than 17 CAG repeats). Finally, the most pronounced difference in AO between sexes was observed among 1273 patients combining the above mentioned stratifications—in those with the mutation of 40–45 CAG repeats and normal allele longer than 17 CAG repeats.

Data availability

The REGISTRY data are not publicly available and are granted by the EHDN to qualified researchers given institutional assurance regarding patients' data confidentiality. Anonymized patient data are granted from the EHDN following the standard submission procedure (<https://ehdn.org/hd-clinicians-researchers/data-mining-projects-2/>). Further information and data requests should be directed to Anna Stanisławska-Sachadyn (anna.stanislawski@pg.edu.pl).

Received: 22 December 2023; Accepted: 5 June 2024

Published online: 08 July 2024

References

- Medina, A., Mahjoub, Y., Shaver, L. & Pringsheim, T. Prevalence and incidence of Huntington's disease: An updated systematic review and meta-analysis. *Mov. Disord.* **37**, 2327–2335. <https://doi.org/10.1002/mds.29228> (2022).
- Ross, C. A. & Tabrizi, S. J. Huntington's disease: From molecular pathogenesis to clinical treatment. *Lancet. Neurol.* **10**, 83–98. [https://doi.org/10.1016/S1474-4422\(10\)70245-3](https://doi.org/10.1016/S1474-4422(10)70245-3) (2011).
- Ross, C. A. *et al.* Huntington disease: Natural history, biomarkers and prospects for therapeutics. *Nat. Rev. Neurol.* **10**, 204–216. <https://doi.org/10.1038/nrneurol.2014.24> (2014).
- Duyao, M. *et al.* Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nat. Genet.* **4**, 387–392. <https://doi.org/10.1038/ng0893-387> (1993).
- Kovtun, I. V., Therneau, T. M. & McMurray, C. T. Gender of the embryo contributes to CAG instability in transgenic mice containing a Huntington's disease gene. *Hum. Mol. Genet.* **9**, 2767–2775. <https://doi.org/10.1093/hmg/9.18.2767> (2000).
- Kovtun, I. V., Welch, G., Guthrie, H. D., Hafner, K. L. & McMurray, C. T. CAG repeat lengths in X- and Y-bearing sperm indicate that gender bias during transmission of Huntington's disease gene is determined in the embryo. *J. Biol. Chem.* **279**, 9389–9391. <https://doi.org/10.1074/jbc.M313080200> (2004).
- Nahhas, F., Garbern, J., Feely, S. & Feldman, G. L. An intergenerational contraction of a fully penetrant Huntington disease allele to a reduced penetrance allele: Interpretation of results and significance for risk assessment and genetic counseling. *Am. J. Med. Genet. Part A* **149a**, 732–736. <https://doi.org/10.1002/ajmg.a.32720> (2009).
- Tang, Y. *et al.* Intergeneration CAG expansion and contraction in a Chinese HD family. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **141b**, 242–244. <https://doi.org/10.1002/ajmg.b.30261> (2006).
- Wheeler, V. C. *et al.* Factors associated with HD CAG repeat instability in Huntington disease. *J. Med. Genet.* **44**, 695–701. <https://doi.org/10.1136/jmg.2007.050930> (2007).
- Foroud, T., Gray, J., Ivashina, J. & Conneally, P. M. Differences in duration of Huntington's disease based on age at onset. *J. Neurol. Neurosurg. Psychiatry* **66**, 52–56. <https://doi.org/10.1136/jnnp.66.1.52> (1999).

11. Genetic_Modifiers_of_Huntington's_Disease_(GeM-HD)_Consortium. CAG repeat not polyglutamine length determines timing of Huntington's disease onset. *Cell* **178**, 887–900.e814. <https://doi.org/10.1016/j.cell.2019.06.036> (2019).
12. Hentosh, S. *et al.* Sex differences in Huntington's disease: Evaluating the enroll-HD database. *Mov. Disord. Clin. Pract.* **8**, 420–426. <https://doi.org/10.1002/mdc3.13178> (2021).
13. Roos, R. A. *et al.* Age at onset in Huntington's disease: Effect of line of inheritance and patient's sex. *J. Med. Genet.* **28**, 515–519. <https://doi.org/10.1136/jmg.28.8.515> (1991).
14. Kennedy, L. *et al.* Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum. Mol. Genet.* **12**, 3359–3367. <https://doi.org/10.1093/hmg/ddg352> (2003).
15. Gomes-Pereira, M. *et al.* Disease-associated CAG-CTG triplet repeats expand rapidly in non-dividing mouse cells, but cell cycle arrest is insufficient to drive expansion. *Nucleic Acids Res.* **42**, 7047–7056. <https://doi.org/10.1093/nar/gku285> (2014).
16. Gonitel, R. *et al.* DNA instability in postmitotic neurons. *Proc. Natl. Acad. Sci. USA* **105**, 3467–3472. <https://doi.org/10.1073/pnas.0800048105> (2008).
17. Swami, M. *et al.* Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum. Mol. Genet.* **18**, 3039–3047. <https://doi.org/10.1093/hmg/ddp242> (2009).
18. Manley, K., Shirley, T. L., Flaherty, L. & Messer, A. Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat. Genet.* **23**, 471–473. <https://doi.org/10.1038/70598> (1999).
19. Pinto, R. M. *et al.* Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: Genome-wide and candidate approaches. *PLoS Genet.* **9**, e1003930. <https://doi.org/10.1371/journal.pgen.1003930> (2013).
20. Genetic_Modifiers_of_Huntington's_Disease_(GeM-HD)_Consortium. Identification of genetic factors that modify clinical onset of Huntington's disease. *Cell* **162**, 516–526. <https://doi.org/10.1016/j.cell.2015.07.003> (2015).
21. Goold, R. *et al.* FAN1 modifies Huntington's disease progression by stabilizing the expanded HTT CAG repeat. *Hum. Mol. Genet.* **28**, 650–661. <https://doi.org/10.1093/hmg/ddy375> (2019).
22. Lee, J. M. *et al.* A modifier of Huntington's disease onset at the MLH1 locus. *Hum. Mol. Genet.* **26**, 3859–3867. <https://doi.org/10.1093/hmg/ddx286> (2017).
23. Iyer, R. R. & Pluciennik, A. DNA mismatch repair and its role in Huntington's disease. *J. Huntingt. Dis.* **10**, 75–94. <https://doi.org/10.3233/jhd-200438> (2021).
24. Kehoe, P., Krawczak, M., Harper, P. S., Owen, M. J. & Jones, A. L. Age of onset in Huntington disease: Sex specific influence of apolipoprotein E genotype and normal CAG repeat length. *J. Med. Genet.* **36**, 108–111 (1999).
25. Snell, R. G. *et al.* Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. *Nat. Genet.* **4**, 393–397. <https://doi.org/10.1038/ng0893-393> (1993).
26. Aziz, N. A., van Belzen, M. J., Coops, I. D., Belfroid, R. D. & Roos, R. A. Parent-of-origin differences of mutant HTT CAG repeat instability in Huntington's disease. *Eur. J. Med. Genet.* **54**, e413–418. <https://doi.org/10.1016/j.ejmg.2011.04.002> (2011).
27. Aziz, N. A. *et al.* Normal and mutant HTT interact to affect clinical severity and progression in Huntington disease. *Neurology* **73**, 1280–1285. <https://doi.org/10.1212/WNL.0b013e3181bd1121> (2009).
28. Djoussé, L. *et al.* Interaction of normal and expanded CAG repeat sizes influences age at onset of Huntington disease. *Am. J. Med. Genet. Part A* **119a**, 279–282. <https://doi.org/10.1002/ajmg.a.20190> (2003).
29. Lee, J. M. *et al.* CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. *Neurology* **78**, 690–695. <https://doi.org/10.1212/WNL.0b013e318249f683> (2012).
30. Langbehn, D. R., Brinkman, R. R., Falush, D., Paulsen, J. S. & Hayden, M. R. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. *Clin. Genet.* **65**, 267–277. <https://doi.org/10.1111/j.1399-0004.2004.00241.x> (2004).
31. Langsrud, Ø. ANOVA for unbalanced data: Use type II instead of type III sums of squares. *Stat. Comput.* **13**, 163–167. <https://doi.org/10.1023/a:1023260610025> (2003).
32. Saft, C. *et al.* Apolipoprotein E genotypes do not influence the age of onset in Huntington's disease. *J. Neurol. Neurosurg. Psychiatry* **75**, 1692–1696. <https://doi.org/10.1136/jnnp.2003.022756> (2004).
33. Shelbourne, P. F. *et al.* Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. *Hum. Mol. Genet.* **16**, 1133–1142. <https://doi.org/10.1093/hmg/ddm054> (2007).
34. Telenius, H. *et al.* Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat. Genet.* **6**, 409–414. <https://doi.org/10.1038/ng0494-409> (1994).
35. Larson, E., Fyfe, L., Morton, A. J. & Monckton, D. G. Age-, tissue- and length-dependent bidirectional somatic CAG-CTG repeat instability in an allelic series of R6/2 Huntington disease mice. *Neurobiol. Dis.* **76**, 98–111. <https://doi.org/10.1016/j.nbd.2015.01.004> (2015).
36. Hong, E. P. *et al.* Association analysis of chromosome X to identify genetic modifiers of Huntington's disease. *J. Huntingt. Dis.* <https://doi.org/10.3233/jhd-210485> (2021).
37. Bode, F. J. *et al.* Sex differences in a transgenic rat model of Huntington's disease: Decreased 17beta-estradiol levels correlate with reduced numbers of DARPP32+ neurons in males. *Hum. Mol. Genet.* **17**, 2595–2609. <https://doi.org/10.1093/hmg/ddn159> (2008).
38. Nuzzo, M. T. *et al.* Huntingtin polyQ mutation impairs the 17β-estradiol/neuroglobin pathway devoted to neuron survival. *Mol. Neurobiol.* **54**, 6634–6646. <https://doi.org/10.1007/s12035-016-0337-x> (2017).
39. Jedrak, P. *et al.* Mitochondrial DNA levels in Huntington disease leukocytes and dermal fibroblasts. *Metab. Brain Dis.* **32**, 1237–1247. <https://doi.org/10.1007/s11011-017-0026-0> (2017).
40. DiFiglia, M. *et al.* Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* **277**, 1990–1993. <https://doi.org/10.1126/science.277.5334.1990> (1997).
41. Farrer, L. A. *et al.* The normal Huntington disease (HD) allele, or a closely linked gene, influences age at onset of HD. *Am. J. Hum. Genet.* **53**, 125–130 (1993).

Acknowledgements

We would like to thank all EHDN REGISTRY Study Group Investigators who collected the data and all participating REGISTRY patients. The list of REGISTRY investigators of the European Huntington's Disease Network is included in the Supplementary Information.

Author contributions

A. S-S. designed and conceptualized study, organized data, designed and executed statistical analyses, wrote first paper draft; M. K. performed the statistical analyses; D. Z. interpreted results, designed and critiqued statistical analyses, substantively revised the work; M. Kry. reviewed and critiqued the manuscript; E. Z. designed and critiqued statistical analyses, interpreted results, revised the work and edited it; J. S. designed and conceptualized study, substantively revised the work; and J. L. substantively revised the work. All authors read and approved the final manuscript.

Funding

Supported by data mining project 636 from the European Huntington Disease Network (EHDN) and CHDI Foundation, Inc. The funders had no role in data analysis, decision to publish, or manuscript preparation.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-64105-5>.

Correspondence and requests for materials should be addressed to A.S.-S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

REGISTRY investigators of the European Huntington's Disease Network (EHDN)

Daniel Zielonka⁵ & Jarosław Sławek^{8,9}