Am. J. Hum. Genet. 60:1202–1210, 1997

The Likelihood of Being Affected with Huntington Disease by a Particular Age, for a Specific CAG Size

R. R. Brinkman,[1] M. M. Mezei,[1] J. Theilmann,[1] E. Almqvist,[1,3] and M. R. Hayden[1]

[1]Department of Medical Genetics and [2]Department of Medicine, Division of Neurology, University of British Columbia, Vancouver; and [3]Departments of Molecular Medicine and Clinical Neuroscience, Division of Geriatric Medicine, Karolinska Institute, Stockholm

Summary

Prior studies describing the relationship between CAG size and the age at onset of Huntington disease (HD) have focused on affected persons. To further define the relationship between CAG repeat size and age at onset of HD, we now have analyzed a large cohort of affected and asymptomatic at-risk persons with CAG expansion. This cohort numbered 1,049 persons, including 321 at-risk and 728 affected individuals with a CAG size of 29–121 repeats. Kaplan-Meier analysis has provided curves for determining the likelihood of onset at a given age, for each CAG repeat length in the 39–50 range. The curves were significantly different ($P < .0005$), with relatively narrow 95% confidence intervals (95% CI) (±10%). Penetrance of the mutation for HD also was examined. Although complete penetrance of HD was observed for CAG sizes of ≥42, only a proportion of those with a CAG repeat length of 36–41 showed signs or symptoms of HD within a normal life span. These data provide information concerning the likelihood of being affected, by a specific age, with a particular CAG size, and they may be useful in predictive-testing programs and for the design of clinical trials for persons at increased risk for HD.

Introduction

Huntington disease (HD) is a progressive, neurodegenerative disorder that presents with motor disturbances, psychiatric symptoms, and cognitive decline (Hayden 1981; Harper 1991). Although the mean age at onset is ~40 years of age, 5%–10% of cases have a juvenile onset before age 20 years, whereas late onset (at >50 years of age) occurs in ~20% (Myers et al. 1985). The disease is inexorably progressive, with a duration of ~15–20 years (Hayden 1981; Harper 1991).

Received November 27, 1996; accepted for publication February 21, 1997.

Address for correspondence and reprints: Dr. Michael R. Hayden, Department of Medical Genetics, University of British Columbia, 416-2125 East Mall, NCE Building, Vancouver, B.C. V6T 1Z4 Canada. E-mail: mrh@ulam.generes.ca

0002-9297/97/6005-0022$02.00

The mutation associated with clinical manifestations of the disease is an unstable CAG trinucleotide expansion in exon 1 of a novel gene located on 4p16.3 (Huntington's Disease Collaborative Research Group 1993). Whereas unaffected persons have a range of 6–35 CAG repeats, with 99% of such individuals having <30 repeats, persons affected with HD have CAG sizes of 36–121 repeats (Kremer et al. 1994; Rubinsztein et al. 1996).

In 1994 The Huntington Study Group, a consortium of investigators and practitioners across North America, defined the ranges of CAG repeat length for normal persons as being <30, whereas a CAG of ≥38 indicated a high probability for development of HD (Huntington Study Group 1995). A repeat size of 30–37 was considered "indeterminate" as to whether the patient would develop HD at some time in the future. These guidelines were put into place pending more definitive information concerning the relationship between CAG length and clinical phenotype of the disease. There have been various interpretations of the range of CAG repeat sizes on normal and HD chromosomes (Andrew et al. 1993; Barron et al. 1993; Craufurd and Dodge 1993; De Rooij et al. 1993; Duyao et al. 1993; Goldberg et al.1993*a;* Huntington Disease Collaborative Group 1993; MacMillan et al. 1993; Norremolle et al. 1993; Rubinsztein et al. 1993; Snell et al. 1993; Stine et al. 1993; Zühlke et al. 1993; Benitez et al. 1994; Legius et al. 1994; Trottier et al. 1994; Claes et al. 1995; Soong and Wang 1995), because of different techniques in measuring CAG repeat length, leading to difficulty in comparing results between studies. Originally the DNA analysis consisted of PCR measurement of the CAG repeat, which included the CAG and an adjacent CCG repeat (Goldberg et al. 1993*b;* Huntington's Disease Collaborative Group 1993; Valdes et al. 1993). However, since the CCG-repeat region is polymorphic (Rubinsztein et al. 1993; Andrew et al. 1994), a more accurate PCR technique, which measures only the CAG repeat, has been developed and is the preferred technique for predictive and diagnostic testing (Riess et al. 1993; Warner et al. 1993; Kremer et al. 1995). This has allowed clarification of the ranges of CAG repeat length in persons affected with HD. An analysis of >1,100 affected persons clearly has shown that HD has not occurred with a CAG length <36 repeats (Kremer et al. 1994; Rubinsztein et al. 1996).

However, for patients who are informed that they have inherited the CAG repeat expansion in the HD range (>35), the question often changes from whether they will develop HD to when the disease will manifest. Numerous studies have described a significant inverse relationship between CAG repeat size and age at onset (Andrew et al. 1993; Barron et al. 1993; Craufurd and Dodge 1993; De Rooij et al. 1993; Duyao et al. 1993; Norremolle et al. 1993; Rubinsztein et al. 1993; Simpson et al. 1993; Snell et al. 1993; Stine et al. 1993; Telenius et al. 1993; Zühlke et al. 1993; Ashizawa et al. 1994; Kieburtz et al. 1994; Legius et al. 1994; Novelletto et al. 1994; Trottier et al. 1994; Lucotte et al. 1995; Masuda et al. 1995; Ranen et al. 1995; Yapijakis et al. 1995; Brandt et al. 1996), with the CAG repeat length being invoked to account for ~50% of variation in age at onset (Andrew et al. 1993; Stine et al. 1993; Legius et al. 1994). However, these prior studies of the relationship between age at onset and CAG repeat size have not included asymptomatic, at-risk individuals with CAG length in the affected range. This makes it impossible to take into account the number of asymptomatic persons with a particular CAG at a specific age, and thus it prevents a complete understanding of the relationship between CAG size and age at onset of HD. Both to further explore the relationship between CAG repeat size and the age at onset of clinical manifestations of HD and to determine the risk of developing HD by a certain age in a patient with a particular CAG repeat size, we analyzed CAG size in large numbers of affected and asymptomatic at-risk individuals. Strict criteria were used to determine the age at onset by incorporation of careful clinical follow-up of patients.

Inclusion of the asymptomatic at-risk individuals with expanded CAG repeat lengths, along with use of the more accurate CAG repeat measurement, allowed us to develop estimates of the probability of developing HD, by a particular age, with a given CAG size. The large numbers of individuals in our database also enabled us to clarify the ranges of normal and HD-associated CAG repeat sizes.

HD previously has been considered to be 100% penetrant (Huntington 1872; Hayden 1981; Harper 1991). However, recent data (Nance 1996; Rubinsztein et al. 1996) suggest that, on rare occasions, HD may indeed be not fully penetrant. This data set also has allowed us to study the penetrance of HD at different CAG sizes and to clearly show that penetrance is modified by CAG size.

Subjects and Methods

Subjects

DNA samples from persons with the diagnosis of HD, their at-risk family members, and unrelated controls have been collected in our laboratory since 1984. To date, we have collected information on 4,934 individuals from 1,193 families. A total of 1,593 individuals in our database are affected with HD, and 2,244 are asymptomatic but at risk (i.e., first- or second-degree relatives of an affected individual). Families from Canada and many parts of the world, including families of European, Asian, black South African, Arab, and South American origin, are represented. Samples from all family members of affected individuals were recruited actively, especially asymptomatic and older first- and second-degree relatives of affected individuals, as part of the predictive-testing program at the University of British Columbia clinical genetics clinic.

For the purpose of this study we used those individuals with CAG expansions of the upper allele that were >28, comprising 728 affected and 321 asymptomatic at-risk individuals from 473 families, whose age at onset or oldest age while still asymptomatic could be ascertained. An accurate assessment of the age at onset was performed through both a retrospective review of patient charts and telephone interviews with patients, family members, genetic counselors, and physicians. Age at onset was defined as the first time at which a patient had either neurological or psychiatric symptoms that represented a permanent change from the normal state. The age used for analysis of all asymptomatic individuals was the oldest age when his or her clinical status was last directly confirmed, either at the genetics clinic in Vancouver or by the local, attending physician. Particular attention was paid to confirmation of current age and clinical status of all asymptomatic, at-risk individuals in the HD database who were >65 years of age.

DNA Analysis and Assessment of CAG Repeats

DNA was extracted from leukocytes by standard procedures (Kunkel et al. 1977). The CAG repeat was assessed for all samples by exclusion of the CCG repeat, by use of PCR analysis with primers that flanked either only the CAG repeat (Kremer et al. 1995) or both the CAG repeat and the CCG repeat (Goldberg et al. 1993b), followed by analysis with primers that flank only the CCG repeat (Andrew et al. 1994). Phasing of the CCG repeat was performed by pedigree studies, when necessary. The CAG repeat size was assessed by comparison with sequenced clones of known CAG size.

Data Analysis

For each year of age and CAG size, all individuals, including both those asymptomatic at-risk and those affected, were taken into account to calculate the cumulative probability of having onset of HD by that age, by use of Kaplan-Meier survival analysis (Splus software version 3.1 release 1, 1992; AT&T). Since we had no affected persons with CAG length <36 repeats, individuals with a CAG repeat length >35 were considered to be a cohort at risk, from birth to either neurological or

Table 1

Distribution of Affected Individuals and Asymptomatic Individuals at Risk for HD Who Have CAG Repeat Size >28

Risk Status	Nos. and Ages for CAG Repeat Size of																							
	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51–121	Overall
At risk:																								
No. of individuals	1	8	8	11	18	9	31	12	4	9	13	47	24	31	24	28	13	5	9	4	4	3	5	321
Average current age (years)	69	42	50	38	41	45	49	50	41	36	41	43	41	39	33	32	29	27	28	27	25	26	19	39
Minimum asymptomatic age (years)	69	18	22	11	6	19	7	9	29	10	12	14	21	18	22	13	16	23	18	19	20	23	16	6
Maximum asymptomatic age (years)	69	79	83	53	80	78	93	85	54	66	79	78	75	57	52	52	40	33	40	39	33	29	23	93
SD of current age (years)	NA	22	17	12	20	20	23	23	12	18	22	14	14	11	8	8	7	4	8	9	6	3	3	16
Affected:																								
No. of individuals	0	0	0	0	0	0	0	1	4	2	8	64	74	98	92	95	63	58	39	31	26	13	60	728
Average age at onset (years)								65	57	60	61	56	53	48	43	41	37	36	32	31	28	26	20	41
Minimum age at onset (years)								65	53	35	39	35	33	29	28	25	23	19	17	17	13	16	4	4
Maximum age at onset (years)								65	60	84	71	84	75	65	65	58	48	45	45	45	45	34	35	84
SD of age at onset (years)								NA	3	35	11	10	9	7	8	7	6	5	7	6	9	6	8	13
Total no. of individuals	1	8	8	11	18	9	31	13	8	11	21	111	98	129	116	123	76	63	48	35	30	16	65	1,049

psychiatric onset or until death or last contact (censored). The Kaplan-Meier survival curves were compared by use of log-rank statistics.

There were only 32 individuals with 36–38 CAG repeats, and there were a total of 65 individuals with CAG repeats >50. Therefore, these CAG lengths were excluded from the survival analysis, since the small numbers of individuals at each particular CAG size precluded rigorous statistical analysis.

Results

The distribution of affected and asymptomatic at-risk individuals with CAG repeats >28 is shown in table 1. There were no affected individuals with <36 CAG repeats. The ages of asymptomatic persons with a CAG repeat size of 30–35 are shown in table 2. A total of 963 individuals (728 affected) had CAG repeats >35. Of these, 866 individuals (90%) from 445 families had CAG repeat lengths of 39–50 repeats. The linear association between log age at onset and log CAG size was significant ($P < .001$), with an r^2 value of .73.

The cumulative probability of onset, at 5-year intervals, for a given CAG repeat is shown in table 3, with the complete age distribution for each CAG being shown in figure 1. The mean 95% CI for the probability of age at onset with a given CAG repeat by a particular age was narrow (±10%).

As the CAG repeat length increased from 39 to 50, there was a significant increase in the probability of onset ($P < .0005$) by a given age. For example, whereas an individual with 40 CAG repeats had only a 13% probability of having onset by age 45 years, this increased to 32% for someone with 42 CAG repeats, 73% for 44 repeats, and 100% for a person with 46 CAG repeats (table 3 and fig. 1).

There were no individuals with a CAG repeat length >41 who remained asymptomatic at >56 years of age. This indicated that clinical manifestation of the disease was fully penetrant within a normal life span for this CAG repeat range.

There were, however, several individuals with CAG repeats of 36–41 who had not manifested with symptoms of HD within a normal expected life span (table 4). For example, there were two male individuals, 78 and 85 years of age, with 36 CAG repeats, and there was a male 75 years of age with 39 CAG repeats; all of them were asymptomatic. Furthermore, there was a female with 38 CAG repeats and a male with 40 repeats who were not affected until age 84 years, and there was a male with 41 repeats who was not affected until age 75 years (table 4). All these individuals were of Canadian origin, except for the 78-year-old male with 36 repeats, who was of American origin. However, there

Table 2

Age Distribution of Asymptomatic Individuals with CAG Repeat Size of 30–35

Age (years)	No. of Individuals with CAG Repeat Size of						Total
	30	31	32	33	34	35	
0–19	1	0	1	2	1	3	8
20–29	1	1	2	5	2	4	15
30–39	3	0	1	2	1	4	11
40–49	1	4	6	1	0	5	17
50–59	0	2	1	5	2	7	17
60–64	0	0	0	1	2	1	4
65–69	0	0	0	0	0	0	0
71–74	1	0	0	0	0	2	3
75–79	1	0	0	1	1	1	4
80–84	0	1	0	1	0	1	3
85–89	0	0	0	0	0	2	2
90–95	0	0	0	0	0	1	1
Total	8	8	11	18	9	31	85

Table 3

Cumulative Probability of Onset at Different Ages, for a Given CAG Repeat Size

Age of Subject (years)	Cumulative Probability (95% CI) for CAG Repeat Size of			
	39 (n = 21)	40 (n = 111)	41 (n = 98)	42 (n = 129)
30				.02 (.05–.00)
35		.02 (.05–.00)	.02 (.05–.00)	.05[a] (.09–.01)[a]
40	.07[a] (.20–.00)[a]	.08 (.13–.02)	.12 (.18–.05)	.14 (.20–.07)
45		.13 (.19–.05)	.21 (.30–.12)	.32 (.41–.23)
50	.16 (.33–.00)	.21 (.30–.12)	.38 (.48–.26)	.58 (.66–.47)
55		.36 (.46–.25)	.55 (.64–.42)	.81 (.87–.71)
60		.61 (.70–47)	.77 (.85–.65)	.99 (1.00–.91)
65	.36 (.59–.00)	.80 (.88–.67)	.88[a] (.94–.78)[a]	1.00 (NA)
70	.68 (.87–.20)	.90 (.96–.77)	.94 (.98–.85)	
75	.79[b] (.94–.28)[b]	.95 (.99–.82)	.98 (1.00–.88)	
80				
85		1.00[a] (NA)[a]		
	43 (n = 116)	44 (n = 123)	45 (n = 76)	46 (n = 63)
20				.03 (.07–.00)
25		.01 (.02–.00)	.05 (.10–.00)	.06 (.12–.00)
30	.05 (.10–.01)	.04 (.08–.01)	.17 (.25–.08)	.10 (.17–.02)
35	.18 (.26–.11)	.22 (.30–.14)	.37 (.47–.24)	.41 (.52–.27)
40	.39 (.48–.29)	.49 (.58–.39)	.72 (.81–.59)	.86 (.93–.73)
45	.56 (.65–.45)	.73 (.80–.62)	.91 (.96–.80)	1.00 (NA)
50	.87 (.93–.78)	.89 (.94–.80)	1.00[a] (NA)[a]	
55	.93 (.97–.85)	.96 (.99–.88)		
60	.99[a] (1.00–.91)[a]	1.00[a] (NA)[a]		
65	1.00 (NA)			
	47 (n = 48)	48 (n = 35)	49 (n = 30)	50 (n = 16)
15			.07 (.15–.00)	
20	.04 (.10–.00)	.06 (.13–.00)	.30 (.45–.12)	.19[a] (.36–.00)[a]
25	.16 (.25–.04)	.15 (.26–.02)	.41 (.57–.20)	.39 (.59–.09)
30	.36 (.49–.20)	.46 (.61–.26)	.53 (.68–.30)	.73 (.89–.32)
35	.64 (.76–.45)	.78 (.88–.57)	.77 (.89–.53)	1.00 (NA)
40	.89 (.95–.72)	.89 (.96–.69)	.95[a] (.99–.70)[a]	
45	1.00 (NA)	1.00 (NA)	1.00 (NA)	

[a] Values are for 1 year greater (CAG repeat sizes 42, 43, and 49), 1 year less (CAG repeat sizes 39–41 and 50), or 2 years less (CAG repeat sizes 44 and 45) than the stated interval.

[b] Value is for an individual 71 years of age.

were no instances of nonpenetrance with a CAG length >41 repeats.

A difference of a single CAG repeat length has a significant effect on the expected age at onset for an individual. There was a significant linear trend between CAG repeat length and median age at onset ($P < .001$; $r^2 = .96$), with the median age at onset decreasing by 3.4 years (±0.2) for each CAG length increase in the 39–50 range, as shown in table 5. For example, although only 50% of persons with 40 CAG repeats will be affected by age 59 years, this decreases to age 37 years for 45 CAG repeats and to age 27 years for 50 CAG repeats.

To assess the effect of any possible bias introduced by the inclusion of multiple individuals from 445 families, we randomly selected only two individuals and repeated the analysis. There was no significant difference in the results obtained for 39-49 CAG repeats, suggesting that we did not introduce any obvious bias into our database by including two or more individuals from any one family for this CAG range.

Discussion

Using a large cohort ascertained uniformly and analyzed by use of the same accurate measures for CAG

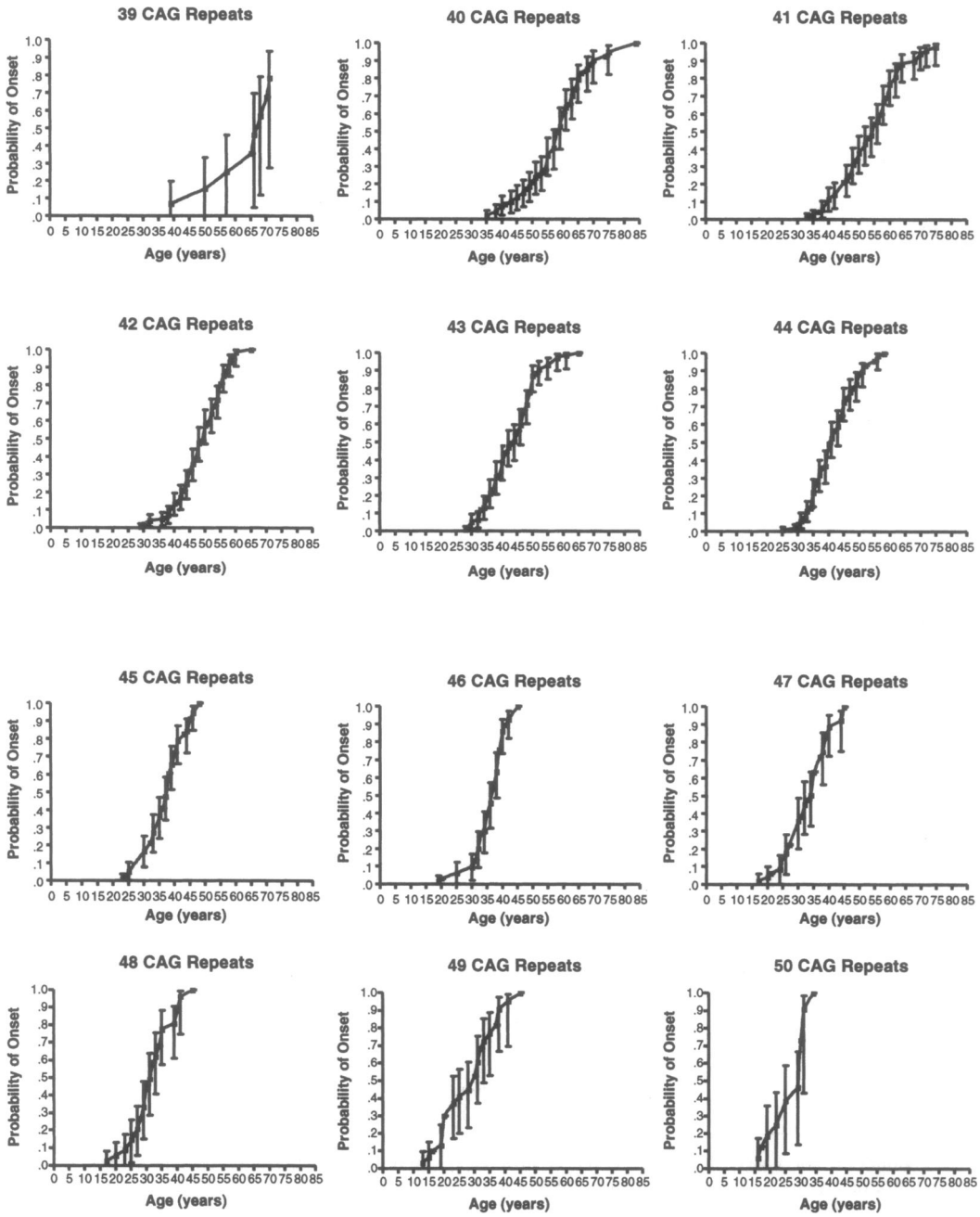

Figure 1 Representations of cumulative probability of being affected, at a particular age, for a CAG of 39–50. Error bars represent 95% CI.

size, this study has clearly shown that CAG repeat length is the major determinant of age at onset in HD. By assessing the CAG size alone, we were able to predict the likelihood that an individual would be affected by a particular age, with relatively narrow 95% CI (±10%) for the vast majority of persons being tested. This study confirms the prior assessments of ranges of CAG repeat length in affected persons and clearly shows that reduced penetrance for HD may occur only for a CAG repeat length <42. The degree of association between CAG size and age at onset was significant ($P < .001$), with an r^2 of .73, which is higher than prior estimates of ~.6 (Andrew et al. 1993; Duyao et al. 1993; Ranen et al. 1995), probably because of definitive determination of CAG size and confirmation of clinical status.

The inverse relationship between the increased CAG repeat size and age at onset for other diseases associated with CAG expansion in novel genes is well documented (La Spada et al. 1992; Jodice et al. 1994; Koide et al. 1994; Nagafuchi et al. 1994; Ranum et al. 1994; Komure et al. 1995; Maciel et al. 1995; Maruyama et al. 1995; Takiyama et al 1995). For HD, since the range

Table 4

Estimation of Penetrance of CAG Expansion in HD Gene, by CAG Repeat Size

CAG Repeat Size	No. of Affected Individuals	No. of Unaffected Individuals, Age >75 Years (Males) or >80 Years (Females)	Total
<29–35	0	9	9
36	1	2	3
37	4	0	4
38	2	1	3[a]
39	8	1	9[a]
40	64	1	65
41	74	1	75
>41	575	0	575

[a] Including individuals reported by Rubinsztein et al. (1996).

of age at onsets for a particular CAG repeat size in the HD gene was very broad, most authors, including ourselves, have recommended against the use of the CAG repeat size to predict the age at onset for an individual patient (Andrew et al. 1993; Barron et al. 1993; Craufurd and Dodge 1993; De Rooij et al. 1993; Duyao et al. 1993; MacMillan et al. 1993; Norremole et al. 1993; Simpson et al. 1993; Stine et al. 1993; Novelletto et al. 1994; Trottier et al. 1994). In this article, including asymptomatic at-risk as well as affected individuals has allowed development of accurate survival curves to predict the probability that an individual will be affected with HD by a certain age. These curves were based on both affected and asymptomatic at-risk individuals with CAG repeat sizes of 39–50, which represents 90% of individuals in our database who have a CAG repeat size in the range of ⩾36. The survival curves were significantly different for each CAG repeat size studied, lending further support to the importance of the CAG repeat size as the predominant factor in determining the age at onset. Although these probability curves cannot be used to predict the particular age at onset for an individual, this analysis may have clinical utility by providing estimates of symptom-free survival to an individual seeking additional information in a predictive-testing program.

Data generated from this study also may have significant implications for the design of clinical trials for new therapeutics. The aim of new therapies in HD is to slow or stop the progression of HD in affected persons and to delay or prevent onset in persons with CAG >35 repeats. An appropriate design of clinical trials for increased-risk individuals will need to take into account the expected age at onset of HD for a particular person, to determine the potential efficacy of therapy. For example, all persons with a CAG repeat length of 46 would be expected to manifest with symptoms by 45 years of age, with a median age at onset of 36 years (see table 3 and fig. 1). Extension of age at onset beyond this age could indicate a significant therapeutic effect of a particular drug. Alternatively, shifting of the median age at onset to >36 years also might indicate a therapeutic effect. These data indicate the importance of having sufficient persons with a particular CAG size in a drug trial, allowing for more rapid ascertainment of a beneficial effect.

The results of this study, however, should be interpreted with caution. Although we used a highly accurate methodology to calculate the CAG repeat size, this analysis cannot be extrapolated immediately to other laboratories, because of possible interlaboratory variability in the PCR assays. However, it is somewhat reassuring in this regard that comparison of CAG repeat lengths on the same samples, between laboratories with significant experience in assessment of CAG size, revealed few differences in assessment of CAG repeat size (Marshall and Huntington Study Group 1996; Rubinsztein et al. 1996).

Another cautionary note is that these results were obtained by use of affected and unaffected persons from families with HD. Therefore, these data may not apply equally to asymptomatic individuals in the general population who have no relative with HD but who are found to have CAG repeats in the range seen in affected persons with HD. However, it is indeed extremely rare to find such individuals in the general population. In addition, it should be recognized that a potential bias in the database is underrepresentation of asymptomatic persons at increased risk for developing HD. This would result in underestimation of the age at onset in this study. However, this analysis did include 661 affected and 205 asymptomatic individuals at risk, and, although this does not represent a complete assessment of CAG in a defined population of affected and at-risk individuals, there has been no systematic bias in ascertainment of data.

Table 5

Median Age at Onset

CAG Repeat Size	Median Age at Onset[a] (95% CI) (years)
39	66 (72–59)
40	59 (61–56)
41	54 (56–52)
42	49 (50–48)
43	44 (45–42)
44	42 (43–40)
45	37 (39–36)
46	36 (37–35)
47	33 (35–31)
48	32 (34–30)
49	28 (32–25)
50	27 (30–24)

[a] Age by which 50% of individuals will be affected.

Our results support the recent findings of Rubinsztein et al. (1996), who investigated individuals with 30–40 repeats. We confirm that the lower limit of CAG repeat size in individuals who manifest with HD is 36. This lower limit is supported both by the fact that there were 31 at-risk individuals with a CAG repeat size of 35, including one man who was 93 years old, and that all these persons were asymptomatic (table 2). It is now justified that the CAG size of 30–35 no longer need be considered part of the indeterminate range. These individuals can be informed that there is no clear documentation of any person manifesting with HD who has CAG repeats in this range. However, the risk that offspring will develop HD may be increased, particularly if the transmitting parent is a male (Goldberg et al. 1995; Chong et al. 1997).

Our database includes five males who were asymptomatic at age ≥75 years and one female with onset at 84 years of age (table 4). Penetrance is defined as the proportion of individuals with a specified genotype who show the expected phenotype under a defined set of environmental conditions, which, in this instance, is a normal expected life span (King et al. 1992). In the past, HD was presumed to be 100% penetrant, with all carriers of the HD expansion manifesting the disease (Huntington 1872; Hayden 1981; Harper 1991). The finding of reduced penetrance has been raised previously elsewhere (Nance 1996; Rubinsztein et al. 1996). Using onset at ≥75 years for males and ≥81 years for females as being beyond the normal life span (Statistics Canada 1995), our analysis of survival curves indicates that there is complete penetrance with a CAG repeat size of ≥42. Furthermore, it is apparent that reduced penetrance may occur within the range of 36–41 CAG repeats. Clearly, these data need validation in other independently ascertained large groups of patients, since the numbers are too small to allow for meaningful penetrance estimates for each specific repeat size. However, it is obvious that there is a trend to increasing penetrance with increasing repeat length in the 36–41-repeat range: <90% for 39 CAG repeats and 99% for 41 CAG repeats (table 4).

The data presented in this article provide significant new information for persons and professionals involved in predictive-testing programs. Accurate genotyping provides assessment as to whether a patient has inherited DNA changes associated with HD, but it also can help to define the likelihood of being affected at a certain age in those individuals with a CAG repeat of 39–50. Furthermore, this information also may be useful for improvement of the design of clinical trials of therapeutic agents for increased-risk individuals.

Acknowledgments

The authors are grateful to Drs. Berry Kremer and Jamal Nasir for helpful comments and to Elizabeth Starr for technical assistance. This article also was strengthened by comments from David Oakes of the Huntington Study Group. The authors also wish to acknowledge all doctors and patients who have contributed blood samples to our HD DNA bank and clinical information to the HD database. This work was supported by the Medical Research Council (Canada) and the Huntington Society of Canada.

References

Andrew SE, Goldberg YP, Kremer B, Telenius H, Theilmann J, Adam S, Starr E, et al (1993) The relationship between trinucleotide repeat (CAG) length and clinical features of Huntington disease. Nat Genet 4:398–403

Andrew SE, Goldberg YP, Theilmann J, Zeisler J, Hayden MR (1994) A CCG repeat polymorphism adjacent to the CAG repeat in the Huntington disease gene: implications for diagnostic accuracy and predictive testing. Hum Mol Genet 3: 65–67

Ashizawa T, Wong LJ, Richards CS, Caskey CT, Jankovic J (1994) CAG repeat size and clinical presentation in Huntington's disease. Neurology 44:1137–1143

Barron L, Warner J, Porteous M, Holloway S, Simpson S, Davidson R, Brock DJH (1993) A study of the Huntington's disease associated trinucleotide repeat in the Scottish population. J Med Genet 30:1003–1007

Benitez J, Fernandez E, Garcia Ruiz P, Robledo M, Ramos C, Yebenes J (1994) Trinucleotide (CAG) repeat expansion in chromosomes of Spanish patients with Huntington's disease. Hum Genet 94:563–564

Brandt J, Bylsma FW, Gross R, Stine OC, Ranen NG, Ross CA (1996) Trinucleotide repeat length and clinical progression in Huntington's disease. Neurology 46:527–531

Chong SS, Almqvist E, Telenius H, LaTray L, Nichol K, Bourdelat-Parks B, Goldberg YP, et al (1997) Contribution of DNA sequence and CAG size to mutation frequencies of intermediate alleles for Huntington disease: evidence from single sperm analyses. Hum Mol Genet 6:301–309

Claes S, Van Zand K, Legius E, Dom R, Malfroid M, Baro F, Godderis J, et al (1995) Correlations between triplet repeat expansion and clinical features in Huntington's disease. Arch Neurol 52:749–753

Craufurd D, Dodge A (1993) Mutation size and age at onset in Huntington's disease. J Med Genet 30:1008–1011

De Rooij KE, De Koning Gans PA, Skraastad MI, Belfroid RD, Vegter-van der Vlis M, Roos RA, Bakker E, et al (1993) Dynamic mutation in Dutch Huntington's disease patients: increased paternal repeat instabilities extending to within the normal size range. J Med Genet 30:996–1002

Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Frontali M, Folstein S, et al (1993) Trinucleotide repeat length instability in Huntington disease. Nat Genet 4:387–392

Goldberg YP, Andrew SE, Clarke LA, Hayden MR (1993*a*) A PCR method for accurate assessment of trinucleotide repeat expansion in Huntington disease. Hum Mol Genet 2:635–636

Goldberg YP, Kremer B, Andrew SE, Theilmann J, Graham RK, Squitieri F, Telenius H, et al (1993*b*) Molecular analysis of new mutations causing Huntington disease: intermediate alleles and sex of origin effects. Nat Genet 5:174–179

Goldberg YP, McMurray CT, Zeisler J, Almqvist E, Sillence D, Richards F, Gacy AM, et al (1995) Increased instability of intermediate alleles in families with sporadic Huntington disease compared to similar sized intermediate alleles in the general population. Hum Mol Genet 4:1911–1918

Harper PS (1991) Huntington's disease. WB Saunders, London

Hayden MR (1981) Huntington's chorea. Springer, London, Berlin, Heidelberg

Huntington G (1872) On chorea. Med Surg Rep 26:317–321

Huntington Study Group (1995) Huntington Study Group (HSG) policy on non-disclosure of CAG trinucleotide repeat length (CAG_n) genetic (DNA). Huntington Study Group, Rochester, NY

Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983

Jodice C, Malaspina P, Persichetti F, Novelletto A, Spadaro M, Giunti P, Morocutti C, et al (1994) Effect of trinucleotide repeat length and parental sex on phenotypic variation in spinocerebellar ataxia 1. Am J Hum Genet 54:959–965

Kieburtz K, MacDonald M, Shih C, Feigin A, Steinberg K, Bordwell K, Zimmerman C, et al (1994) Trinucleotide repeat length and progression of illness in Huntington's disease. J Med Genet 872–874

King RA, Rotter JI, Motulsky AG (1992) The genetic basis of common diseases. Oxf Monogr Med Genet 20:3–18

Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, Takahashi H, et al (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat Genet 6:9–13

Komure O, Sano A, Nishino N, Yamauchi N, Ueno S, Konodoh K, Sano N, et al (1995) DNA analysis in hereditary dentatorubral-pallidoluysian atrophy: correlation between CAG repeat length and phenotypic variation and the molecular basis of anticipation. Neurology 45:143–149

Kremer B, Almqvist E, Theilmann J, Spence N, Telenius H, Goldberg YP, Hayden MR (1995) Sex-dependent mechanisms for expansions and contractions of the CAG repeat on affected Huntington disease chromosomes. Am J Hum Genet 57:343–350

Kremer HPH, Goldberg YP, Andrew SE, Theilmann J, Telenius H, Zeisler J, Squitieri F, et al (1994) A worldwide study of the Huntington's disease mutation: the sensitivity and specificity of measuring CAG repeats. N Engl J Med 330: 1401–1406

Kunkel LM, Smith KD, Boyer SM, Borgaonkar DS, Wachtel SS, Miller OS, Breg WR, et al (1977) Analysis of human Y-chromosome-specific reiterated DNA in chromosome variants. Proc Natl Acad Sci USA 74:1245–1249

La Spada AR, Roling DB, Harding AE, Warner CL, Spiegel R, Hausmanowa-Petrusewicz I, Yee W-C, et al (1992) Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. Nat Genet 2:301–304

Legius E, Cuppens H, Dierick H, Van Zandt K, Dom R, Fryns J-P, Evers-Kiebooms G, et al (1994) Limited expansion of the $(CAG)_n$ repeat of the Huntington gene: a premutation(?). Eur J Hum Genet 2:44–50

Lucotte G, Turpin JC, Riess O, Epplen JT, Siedlaczk I, Loirat F, Hazout S (1995) Confidence intervals for predicted age of onset, given the size of $(CAG)_n$ repeat in Huntington disease. Hum Genet 95:231–232

Maciel P, Gaspar C, DeStefano AL, Silveira I, Coutinho P, Radvany J, Dawson DM, et al (1995) Correlation between CAG repeat length and clinical features in Machado-Joseph disease. Am J Hum Genet 57:54–61

MacMillan JC, Snell RG, Tyler A, Houlihan GD, Fenton I, Cheadle JP, Lazarou LP, et al (1993) Molecular analysis and clinical correlations of the Huntington's disease mutation. Lancet 342:954–958

Marshall FJ, Huntington Study Group (1996) Interlaboratory variability of $(CAG)_n$ determinations in Huntington's disease. Neurology 46:258

Maruyama H, Nakamura S, Matsuyama Z, Sakai T, Doyu M, Sobue G, Seto M, et al (1995) Molecular features of the CAG repeats and clinical manifestation of Machado-Joseph disease. Hum Mol Genet 4:807–812

Masuda N, Goto J, Murayama N, Watanabe M, Kondo I, Kanazawa I (1995) Analysis of triplet repeats in the huntingtin gene in Japanese families affected with Huntington's disease. J Med Genet 32:701–705

Myers RH, Sax DS, Schoenfeld M, Bird ED, Wolf PA, Vonsattel JP, White RF, et al (1985) Late onset of Huntington's disease. J Neurol Neurosurg Psychiatry 48:530–534

Nagafuchi S, Yanagisawa H, Sato K, Shirayama T, Ohsaki E, Bundo M, Takeda T, et al (1994) Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. Nat Genet 6:14–18

Nance MA (1996) Huntington disease—another chapter rewritten. Am J Hum Genet 59:1–6

Norremolle A, Riess O, Epplen JT, Fenger K, Hasholt L, Sorenson SA (1993) Trinucleotide repeat elongation in the Huntington gene in Huntington disease patients from 71 Danish families. Hum Mol Genet 2:1475–1476

Novelletto A, Persichetti F, Sabbadini G, Mandich P, Bellone E, Ajmar F, Pergola M, et al (1994) Analysis of the trinucleotide repeat expansion in Italian families affected with Huntington disease. Hum Mol Genet 3:93–98

Ranen NG, Stine OC, Abbott MH, Sherr M, Codori A-M, Franz ML, Chao NI, et al (1995) Anticipation and instability of IT-15 $(CAG)_N$ repeats in parent-offspring pairs with Huntington disease. Am J Hum Genet 57:593–602

Ranum LPW, Chung M-y, Banfi S, Bryer A, Schut LJ, Ramesar R, Duvick LA, et al (1994) Molecular and clinical correlations in spinocerebellar ataxia type 1: evidence for familial effects on the age at onset. Am J Hum Genet 55:244–252

Riess O, Norremoelle A, Soerensen SA, Epplen JT (1993) Improved PCR conditions for the stretch of $(CAG)_n$ repeats causing Huntington's disease. Hum Mol Genet 2:637

Rubinsztein DC, Barton DE, Davison BCC, Ferguson-Smith MA (1993) Analysis of the huntingtin gene reveals a trinucleotide-length polymorphism in the region of the gene that contains two CCG-rich stretches and a correlation between decreased age of onset of Huntington's disease and CAG repeat number. Hum Mol Genet 2:1713–1715

Rubinsztein DC, Leggo J, Coles R, Almqvist E, Biancalana V, Cassiman J-J, Chotai K, et al (1996) Phenotypic characterization of individuals with 30–40 CAG repeats in the Huntington disease gene (HD) reveals HD cases with 36 repeats and apparently normal elderly individuals with 36–39 repeats. Am J Hum Genet 59:16–22

Simpson SA, Davidson MJ, Barron LH (1993) Huntington's disease in Grampian region: correlation of the CAG repeat number and the age of onset of the disease. J Med Genet 30:1014–1017

Snell RG, MacMillan JC, Cheadle JP, Fenton I, Lazarou LP, Davies P, MacDonald ME, et al (1993) Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. Nat Genet 4:393–397

Soong B-W, Wang JT (1995) A comparison of the Huntington's disease associated trinucleotide repeat between Chinese and white populations. J Med Genet 32:404–408

Statistics Canada (1995) Life tables, Canada and provinces, 1990–1992. Catalogue 84-537. Statistics Canada, Ottawa

Stine OC, Pleasant N, Franz ML, Abbott MH, Folstein SE, Ross CA (1993) Correlation between the onset age of Huntington's disease and length of the trinucleotide repeat in IT-15. Hum Mol Genet 2:1547–1549

Takiyama Y, Igarashi S, Rogaeva EA, Endo K, Rogaev EI, Tanaka H, Sherrington R, et al (1995) Evidence for intergenerational instability in the CAG repeat in the MJD1 gene and for conserved haplotypes at flanking markers amongst Japanese and Caucasian subjects with Machado-Joseph disease. Hum Mol Genet 4:1137–1146

Telenius H, Kremer HPH, Theilmann J, Andrew SE, Almqvist E, Anvret M, Greenberg C, et al (1993) Molecular analysis of juvenile Huntington disease: the major influence on $(CAG)_n$ repeat length is the sex of the affected parent. Hum Mol Genet 2:1535–1540

Trottier Y, Biancalana V, Mandel JL (1994) Instability of CAG repeats in Huntington's disease: relationship to parental transmission and age of onset. J Med Genet 31:377–382

Valdes JM, Tagle DA, Elmer LW, Collins FS (1993) A simple non-radioactive method for diagnosis of Huntington's disease. Hum Mol Genet 2:633–634

Warner JP, Barron L, Brock DJP (1993) A new polymerase chain reaction (PCR) assay for the trinucleotide repeat that is unstable and expanded in Huntington's disease. Mol Cell Probes 7:235–239

Yapijakis C, Vasselopoulos D, Tzagournisakis M, Maris T, Fesdjian C, Papageorgion C, Plaitakis A (1995) Linkage disequilibrium between the expanded $(CAG)_n$ repeat and an allele of the adjacent $(CCG)_n$ repeat in Huntington's disease patients of Greek origin. Eur J Hum Genet 3:228–234

Zühlke C, Riess O, Shroder K, Siedlaczck I, Epplen JT, Engel W, Thies U (1993) Expansion of the $(CAG)_n$ repeat causing Huntington's disease in 352 patients of German origin. Hum Mol Genet 2:1467–1469