Short Report

Estimation of the age at onset in spinocerebellar ataxia type 2 Cuban patients by survival analysis


Previous studies have investigated the close association that exists between CAG repeat number and the age at onset in SCA2 = spinocerebellar ataxia type 2. These studies have focused on affected individuals. To further characterize this association and estimate the risk of a carrier developing SCA2 at a particular age as a function of a specific CAG repeat size, we have analyzed a large group of 924 individuals, including 394 presymptomatic and 530 affected individuals with a CAG repeat length of 32–79 units. Using a Kaplan–Meier survival analysis, we obtained cumulative probability curves for disease manifestation at a particular age for each CAG repeat length in the 34–45 range. These curves were significantly different (p < 0.001) and showed small overlap. All these information may be very valuable in predictive-testing programs, in the planning of studies for the identification of other genetic and environmental factors as modifiers of age at onset, and in the design of clinical trials for people at enlarged risk for SCA2.

LE Almaguer-Mederos, NS Falcón, YR Almira, YG Zaldivar, DC Almarales, EM Góngora, MP Herrera, KE Batallán, RR Armiñán, MV Manresa, GS Cruz, J Laffita-Mesa, TM Cyuz, V Chang, G Auburger, S Gispert and LV Pérez

aCenter for Investigation and Rehabilitation of Hereditary Ataxias (CIRAH), Holguín Province, Cuba, and
bSection of Molecular Neurogenetics, Department of Neurology, J.W. Goethe University Medical School, Building 89, 3rd Floor, Theodor Stern Kai 7, 60590 Frankfurt am Main, Germany

Key words: age at onset – polyglutamine disorders – spinocerebellar ataxia type 2 – survival analysis

Corresponding author: Luis Enrique Almaguer-Mederos, Center for Investigation and Rehabilitation of Hereditary Ataxias (CIRAH), Edif. 7, Apto. 17, Nuevo Holguín Holguín, CP 80100 Cuba. Tel.: +53 24 424090; fax: +53 24 463579; e-mail: leam@cristal.hlg.sld.cu

Received 15 September 2009, revised and accepted for publication 30 November 2009

Spinocerebellar ataxia type 2 (SCA2) is a progressive, neurodegenerative disorder consisting of gait ataxia accompanied by leg cramps, kinetic, or postural tremor, decreased tendon reflexes and abnormal eye movements with slowed saccades progressing to nuclear ophthalmoplegia. It is due to a CAG repeat expansion mutation in the SCA2 gene coding for a polyglutamine protein -ataxin-2- involved in RNA metabolism and translational regulation (1). Unaffected individuals have 13–31 CAG repeats, whereas affected individuals have 32–79, with some in the range of 500 repeats (2, 3).

The age at onset is highly variable in and between SCA2 families, ranging from 2 to 68 years; 25% of the cases became symptomatic before 22 years (juvenile onset) (4). There is a strong inverse correlation between age at onset and
Almaguer-Mederos et al.

CAG repeat number in SCA2 expanded alleles; however, there is a large variation in age at onset for particular CAG repeat numbers (5) hampering the obtaining of accurate predictive models.

Here, we make use of a non-parametric survival approach to further explore the relationship between age at onset and CAG repeat number in SCA2 expanded alleles, and to estimate the risk of developing SCA2 at a definite age by a particular CAG repeat size. The use of a large sample of affected and presymptomatic SCA2 individuals from the largest and most genetically uniform SCA2 population worldwide allowed us to obtain more accurate figures.

Methods

Subjects

Since 1989, genealogical, clinical, and molecular data from individuals with clinical and/or molecular or presymptomatic diagnosis of SCA2 and relatives have been collected in the Center for the Investigation and Rehabilitation of Hereditary Ataxias (CIRAH) in Holguin, Cuba. Until now, we have collected information on 11,525 individuals from 109 unique SCA2 families. According to the intentions of this study, we assembled a dataset in which we included those heterozygous individuals with more than 31 CAG repeats at the larger allele, whose age at onset or current age, in the case of presymptomatic individuals, could be established. These data encompass 530 affected individuals and 394 presymptomatic individuals from 101 unique SCA2 families. According to the intentions of this study, we assembled a dataset in which we included those heterozygous individuals with more than 31 CAG repeats at the larger allele, whose age at onset or current age, in the case of presymptomatic individuals, could be established. These data encompass 530 affected individuals and 394 presymptomatic individuals from 101 unique SCA2 families. According to the intentions of this study, we assembled a dataset in which we included those heterozygous individuals with more than 31 CAG repeats at the larger allele, whose age at onset or current age, in the case of presymptomatic individuals, could be established. These data encompass 530 affected individuals and 394 presymptomatic individuals from 101 unique SCA2 families. According to the intentions of this study, we assembled a dataset in which we included those heterozygous individuals with more than 31 CAG repeats at the larger allele, whose age at onset or current age, in the case of presymptomatic individuals, could be established. These data encompass 530 affected individuals and 394 presymptomatic individuals from 101 unique SCA2 families. According to the intentions of this study, we assembled a dataset in which we included those heterozygous individuals with more than 31 CAG repeats at the larger allele, whose age at onset or current age, in the case of presymptomatic individuals, could be established. These data encompass 530 affected individuals and 394 presymptomatic individuals from 101 unique SCA2 families. According to the intentions of this study, we assembled a dataset in which we included those heterozygous individuals with more than 31 CAG repeats at the larger allele, whose age at onset or current age, in the case of presymptomatic individuals, could be established. These data encompass 530 affected individuals and 394 presymptomatic individuals from 101 unique SCA2 families. According to the intentions of this study, we assembled a dataset in which we included those heterozygous individuals with more than 31 CAG repeats at the larger allele, whose age at onset or current age, in the case of presymptomatic individuals, could be established. These data encompass 530 affected individuals and 394 presymptomatic individuals from 101 unique SCA2 families. According to the intentions of this study, we assembled a dataset in which we included those heterozygous individuals with more than 31 CAG repeats at the larger allele, whose age at onset or current age, in the case of presymptomatic individuals, could be established. These data encompass 530 affected individuals and 394 presymptomatic individuals from 101 unique SCA2 families. According to the intentions of this study, we assembled a dataset in which we included those heterozygous individuals with more than 31 CAG repeats at the larger allele, whose age at onset or current age, in the case of presymptomatic individuals, could be established. These data encompass 530 affected individuals and 394 presymptomatic individuals from 101 unique SCA2 families.

CAG repeat size determination

Genomic DNA was isolated from peripheral blood leucocytes using a standard protocol (6). The SCA2 CAG repeat was assessed by polymerase chain reaction (PCR) amplification with the previously published UH10 and UH13 oligonucleotide primers (7). Aliquots of the PCR products were then precisely sized by fragment analysis on an ALF Express II apparatus (Amersham Biosciences, Sweden) in comparison to internal and external size markers using Allele Links 1.0 software (Amersham Biosciences, Sweden). Cases with 32 or more repeats were designated SCA2 gene carriers in accordance with published association with disease expression (2).

Statistical analysis

Descriptive statistics were used to describe the central tendencies and dispersion of the variables under study. The Student’s t test was used to look for significant differences between sexes in affected and presymptomatic individuals, regarding the age at onset and current age, respectively. A simple linear regression was used to fit the association between the age at onset and the CAG repeat number in affected individuals. Almost all presymptomatic and affected individuals were considered to calculate the cumulative probability of having onset of SCA2 by a particular age, given a specific CAG repeat expansion, by using Kaplan–Meier survival analysis in SPSS software (version 15.0) (8). The cases with CAG repeat expansions of 32–33 and those with expansions higher than 45 units were excluded from the survival analysis, given the small quantity of individuals at these particular CAG repeat lengths; there were no more than 7 individuals with 32–33 CAG repeats, and a total of 44 persons with CAG repeats of 45–79. Then, 873 individuals (94.5%) with 34–45 CAG repeats were used for Kaplan–Meier survival analysis. Presymptomatic individuals were included in the survival analysis as censored cases. The Kaplan–Meier curves were compared by the use of log-rank statistics. Finally, Spearman’s correlation coefficient was used to test for associations between the CAG repeat numbers, and mean and quartile values resulting from survival analysis for age at onset.

Results

The descriptive statistics for CAG repeat number and current age or age at onset in presymptomatic and affected SCA2 individuals is shown in Table 1. In this cohort, the age at onset showed an asymmetrical distribution, with a skewness (SE) of 0.35 (0.11) and a kurtosis (SE) of −0.49 (0.21), indicating a slightly displaced to the left and a more flatted distribution than a normal one. Non-normality of the age at onset distribution was confirmed by Kolmogorov-Smirnov test (p < 0.001). There was no significant difference between male and female presymptomatics for current age
Discussion

A significant inverse correlation has been established between age at onset and CAG repeat length in SCA2 (2, 5, 9) and other polyglutamine disorders (10, 11). Despite the strength of these associations, there is a remarkable variability of age at onset within each CAG repeat size and, consequently, has been generally recommended not to use the CAG repeat number to make predictions for age at onset on a particular patient (12).

The present investigation, carried out in the largest and most genetically uniform SCA2 population worldwide, makes truly evident that CAG repeat number on expanded alleles is the main determinant of age at onset in SCA2. The degree of association between CAG repeat number and age at onset was highly significant \( r^2 = 0.60; \ p < 0.001 \), which is in range with previous estimates ranging from 0.47 to 0.80 (2, 9, 13, 14). The inclusion of both affected and presymptomatic SCA2 individuals with CAG repeat expansions from 34 to 45 units in this study, led us to apply less-biased statistical techniques that allowed us to obtain survival curves for the prediction of the age at onset for SCA2, with more constricted CI. These curves were significantly different for each studied CAG repeat number, stressing the importance of expanded allele CAG repeat number as the principal factor in determining age at onset in SCA2. The probabilistic model, we have developed, has potential applications for genetic counseling of presymptomatic individuals. Nonetheless, because we cannot exclude the possible occurrence of observation biases and a significant impact of

---

Table 1. Descriptive statistics for CAG repeat number and current age/age at onset for presymptomatic and affected individuals for SCA2

<table>
<thead>
<tr>
<th>Clinical status</th>
<th>N</th>
<th>Range</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presymptomatic</td>
<td>394</td>
<td>3–78</td>
<td>35.1</td>
<td>13.66</td>
<td>32–48</td>
<td>37.9</td>
<td>2.88</td>
</tr>
<tr>
<td>Male</td>
<td>174</td>
<td>3–78</td>
<td>34.8</td>
<td>13.31</td>
<td>33–79</td>
<td>38.3</td>
<td>4.84</td>
</tr>
<tr>
<td>Female</td>
<td>220</td>
<td>7–70</td>
<td>35.4</td>
<td>13.95</td>
<td>32–47</td>
<td>37.8</td>
<td>2.91</td>
</tr>
<tr>
<td>Male</td>
<td>530</td>
<td>2–68</td>
<td>33.1</td>
<td>14.03</td>
<td>37–79</td>
<td>40.1</td>
<td>4.38</td>
</tr>
<tr>
<td>Female</td>
<td>284</td>
<td>2–68</td>
<td>32.5</td>
<td>14.48</td>
<td>32–48</td>
<td>38.3</td>
<td>2.83</td>
</tr>
<tr>
<td>Male</td>
<td>246</td>
<td>5–68</td>
<td>33.7</td>
<td>13.49</td>
<td>32–58</td>
<td>39.9</td>
<td>3.79</td>
</tr>
<tr>
<td>Female</td>
<td>246</td>
<td>5–68</td>
<td>33.7</td>
<td>13.49</td>
<td>32–58</td>
<td>39.9</td>
<td>3.79</td>
</tr>
</tbody>
</table>

\( t = -0.37; \ p = 0.71 \) or for CAG repeat number \( t = 1.77; \ p = 0.078 \). Similarly, there was no significant difference between male and female affected individuals for age at onset \( t = -0.95; \ p = 0.34 \) or for CAG repeat number \( t = 0.73; \ p = 0.47 \). In affected individuals, there was a highly significant linear correlation between age at onset and CAG repeat number \( r = -0.66; \ p < 0.001 \). Logarithmic transformation of the age at onset led to a better prediction \( r = -0.779; \ p < 0.001 \). The difference for age at onset between male and female remained insignificant after adjustment for CAG repeat number \( t = -0.63; \ p = 0.53 \).

The cumulative probability for disease manifestation at a certain age given a particular CAG repeat number is shown in Figure 1. These curves were significantly different \( p < 0.001 \) with quite constricted 95% confidence intervals (CI), and showed small overlap. It could be noted a significant increase in the probability of manifesting disease for a given age, as the CAG repeat number augmented from 34 to 45 units. For instance, all persons with 36 CAG repeat number had simply a 9% probability of being sick by 35 years of age; this figure increased to 26% for those with 38 CAG repeat units, 70% for 41 CAG repeats, and 92% for individuals with a 43 CAG repeat expansion (Fig. 1).

A highly significant linear tendency existed between mean age at onset estimates and CAG repeat number \( r = -0.98; \ p < 0.001 \). The mean age at onset diminished by 4.15 ± 3.45 years for each increase in the CAG repeat number (Table 2). Similarly, the associations between age at onset and CAG repeat number were increasingly weaker from the first to the third age at onset quartiles: \( Q_1, r = -0.998; \ p < 0.001 \); \( Q_2, r = -0.991; \ p < 0.001 \); \( Q_3, r = -0.84; \ p < 0.001 \). The median age at onset diminished by 4.20 ± 2.44 years for each increase in the CAG repeat number, as shown in Table 2.
Fig. 1. Representation of cumulative probability for disease manifestation at a certain age, given a specific CAG repeat length in the range of 34-45. Error bars represent 95% CI.

Table 2. Mean and quartile estimates for age at onset by CAG units in the 34-45 repeat range for spinocerebellar ataxia type 2

<table>
<thead>
<tr>
<th>CAG repeat number</th>
<th>N</th>
<th>Mean age at onset (95% CI) (years)</th>
<th>Quartiles for age at onset (95% CI) (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Q1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Q2&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>34</td>
<td>36</td>
<td>68.8 (61.9–75.8)</td>
<td>62 (27–97)</td>
</tr>
<tr>
<td>35</td>
<td>57</td>
<td>60.8 (57.9–63.6)</td>
<td>55 (52–58)</td>
</tr>
<tr>
<td>36</td>
<td>97</td>
<td>53.7 (50.3–57.0)</td>
<td>49 (41–57)</td>
</tr>
<tr>
<td>37</td>
<td>130</td>
<td>48.5 (46.6–50.3)</td>
<td>42 (40–44)</td>
</tr>
<tr>
<td>38</td>
<td>128</td>
<td>43.3 (40.7–45.9)</td>
<td>35 (32–38)</td>
</tr>
<tr>
<td>39</td>
<td>107</td>
<td>36.1 (34.1–38.0)</td>
<td>30 (28–32)</td>
</tr>
<tr>
<td>40</td>
<td>90</td>
<td>34.5 (32.3–36.6)</td>
<td>29 (26–32)</td>
</tr>
<tr>
<td>41</td>
<td>88</td>
<td>36.3 (32.3–40.4)</td>
<td>25 (23–27)</td>
</tr>
<tr>
<td>42</td>
<td>54</td>
<td>30.1 (25.8–34.5)</td>
<td>23 (20–26)</td>
</tr>
<tr>
<td>43</td>
<td>51</td>
<td>23.5 (21.2–25.7)</td>
<td>19 (18–20)</td>
</tr>
<tr>
<td>44</td>
<td>21</td>
<td>23.3 (19.6–26.9)</td>
<td>19 (16–22)</td>
</tr>
<tr>
<td>45</td>
<td>16</td>
<td>23.2 (15.8–30.6)</td>
<td>13 (10–16)</td>
</tr>
</tbody>
</table>

ND, not determined.

<sup>a</sup>Age by which 25% of individuals will be affected.

<sup>b</sup>Age by which 50% of individuals will be affected.

<sup>c</sup>Age by which 75% of individuals will be affected.

secondary unidentified genetic and environmental factors on the age at onset in our data, we cannot use this information to foresee any individual’s age at onset with confidence. However, these data may be clinically useful offering expected ranges of onset for mutation carrier individuals requesting additional information in the context of predictive testing programs. Genetic counseling based on these estimates might be helpful as it may assist the patient in making more rational
decisions, regarding family planning, reproduction, financial, and health matters. According to our experience, a large proportion of at-risk individuals with a positive result at predictive testing requests such information and experience a relief associated with the reduction of uncertainty (unpublished data).

In addition to its application for genetic counseling, the survival model we have obtained could be useful for planning studies focused on the identification of novel genetic and/or environmental factors that potentially modify the age at onset. For such studies affected individuals should be selected who became sick sooner or later than the specified median 95% CI for their particular CAG repeat size, and even presymptomatic individuals who remain symptom-free beyond the specified 95% CI for their definite CAG repeat length. A sampling strategy like that would increase the power in association studies (15). For example, there has been reported an association between age at onset for SCA2 and CAG repeat polymorphisms in RAI1 (16, 17) and SCA6 genes (5). The identification of novel genetic and/or environmental factors as modifiers of the age at onset is of great importance for the understanding of the molecular mechanisms involved in the pathologic process, and for the recognition of potential therapeutic targets. In addition, the identification of such modifier factors could be useful for developing better models with increased predictive power. Also, the probabilistic estimates we have obtained for age at onset could be useful for the appropriate design of clinical trials for new therapeutic strategies, helping to define the potential efficacy of therapy in presymptomatic individuals. For instance, delaying of disease onset beyond the expected mean or median age might point to a therapeutic result.

Our estimates should not be extrapolated immediately to other SCA2 populations, because of possible interlaboratory differences in the assessment of the CAG repeat number, and because we based them in individuals belonging to SCA2 families, these data may not apply equally to presymptomatic individuals in the general population (‘sporadic’ cases). Although at least two previous studies minimize the significance of the former hypothetically limiting factor by showing very small interlaboratory differences, mainly by only one CAG repeat unit, in assessment of CAG repeat size at SCA2 locus (18, 19), and that ‘sporadic’ cases positive for SCA2 appears to be extremely rare in the general population (2, 20–22), we considered that these cautionary tales should always be taken into consideration the significant impact of a change by only one CAG repeat unit in the survival curves. Additionally, secondary genetic factors may have specific population distributions, affecting the way they influence the age at onset distributions (23).

On the whole, the figures presented in this article offer important new information for individuals and professionals occupied in predictive testing programs, or engaged in studies for the identification of novel genetic and/or environmental factors as modifiers of age at onset, or in the design of clinical trials of therapeutic strategies for presymptomatic individuals.

Acknowledgements

We are grateful to the affected patients, presymptomatic individuals, and Cuban Ministry of Health for their cooperation with this research. We are grateful to Dr Patrick MacLeod for a critical review of the article.

References

8. SPSS Inc. SPSS for Windows (version 15.0) 2006.
14. Geschwind DH, Perlman S, Figueroa CP, Treiman LJ, Pulst SM. The prevalence and wide clinical spectrum of the spinocerebellar ataxia type 2 trinucleotide repeat in patients
Almaguer-Mederos et al.

with autosomal dominant cerebellar ataxia. Am J Hum Genet
15. Rosenblatt A, Brinkman RR, Liang KY et al. Familial influ-
ence on age of onset among siblings with Huntington dis-
399–403.
16. Hayes S, Turecki G, Brisebois K et al. CAG repeat in RA11
is associated with age at onset variability in spinocerebellar
1753–1758.
17. Chattopadhyay B, Ghosh S, Gangopadhyay PK et al. Modula-
tion of age at onset in Huntington’s disease and spinocerebellar
ataxia type 2 patients originated from eastern India. Neurosci
18. Potter NT, Nance MA, and the Ataxia Molecular Diagnostic
19. van de Warrenburg BPC, Sinke RJ, Verschuuren-Bemelmans
CC et al. Spinocerebellar ataxias in the Netherlands:
prevalence and age at onset variance analysis. Neurology 2002:
20. Pujana MA, Corral J, Gratacós M et al. Spinocerebellar atax-
ias in Spanish patients: genetic analysis of familial and spo-
of CAG repeats in SCA1, SCA2, SCA3, and DRPLA loci in
spinocerebellar ataxia patients and distribution of CAG repeats
at the SCA1, in nine ethnic populations of eastern India. Hum
SCA3/MJD, SCA6, SCA7, and DRPLA CAG trinucleotide
repeat expansion in patients with hereditary spinocerebellar
540–544.
23. Andresen JM, Gayán J, Cherny SS et al. Replication of twelve
association studies for Huntington’s disease residual age of
onset in large Venezuelan kindreds. J Med Genet 2007: 44:
44–50.