# HUNTINGTON'S DISEASE AND INSURANCE I: A MODEL OF HUNTINGTON'S DISEASE

By Cristina Gutiérrez and Angus Macdonald

ABSTRACT

We review the literature on the epidemiology of Huntington's disease (HD), a highly penetrant, dominantly inherited, fatal neurological disorder. Although it is a single-gene disorder, mutations are variable in their effects, depending on the number of times that the CAG trinucleotide is repeated in a certain region of the HD gene. Very reliable genetic tests are now available. We fit models: (a) to rates of onset, depending on CAG repeat length as well as age; (b) to post-onset rates of mortality; and (c) to the distribution of CAG repeat lengths in the population. In Part II we use these models to study the critical illness and life insurance markets.

KEYWORDS

CAG Repeat Length; Family History; Genetic Tests; Huntington's Disease; Insurance; Onset Rate; Post-Onset Survival

CONTACT ADDRESS

Angus Macdonald, Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, U.K. Tel: +44(0)131-451-3209; Fax: +44(0)131-451-3249; E-mail: A.S.Macdonald@ma.hw.ac.uk

## 1. Introduction

Early actuarial studies of genetics and insurance, such as Macdonald (1999), used generic models of genetic risk, making broad assumptions about entire classes of genetic disorders rather than modelling individual disorders. These were useful in obtaining certain kinds of 'null' results, for example that even under extreme assumptions, multifactorial disorders were unlikely to be significant in life insurance. However these models do not address the questions that arise in respect of specific disorders. In the United Kingdom, the Genetics and Insurance Committee (GAIC) of the Department of Health has the task of assessing applications from insurers to be allowed to use specific genetic test results in underwriting specific kinds of insurance. This needs 'bottom-up' rather than 'top-down' models of genetic risk (Macdonald, 2003), based on the epidemiology of each disorder. This paper presents such a model, in respect of Huntington's disease (HD). In Part I, we review the epidemiological literature on HD, and propose models for those aspects most relevant to insurance questions. In Part II we apply these models to questions of insurance premiums and the costs of adverse selection.

HD, or Huntington's chorea (from the Greek for 'dance') is a severe, progressive and ultimately fatal form of dementia, usually occurring at ages 30 and over. Patients with HD can survive for 30 years or more after its appearance, but half die of pneumonia, choking or heart failure within 15–20 years. There may be a higher than normal risk of suicide.

HD has long been one of the clearest examples of a late-onset genetic disorder caused by mutations in a single gene, now called the HD gene:

(a) Everyone has two copies of the HD gene, one inherited from each parent. (It is not correct to refer to the HD gene as 'a disease gene'. It is a vital gene in every person's genome, but mutations in the gene may cause disease.)
(b) Certain mutations in the HD gene encode a faulty gene product (a protein called huntingtin) that gradually affects brain cells, leading to HD. Therefore, someone with a mutation in just one of their two copies of the gene is at risk of HD.
(c) The mutation penetrance (that is, the lifetime risk of HD, in the absence of all other risks, for someone with a mutation) is extremely high, practically 100%.
(d) Mutations are rare, so the chances that anyone has mutations in both copies of the HD gene can be ignored (almost equivalent to assuming that it is unlikely that anyone's mother and father each have a mutation).
(e) The combination of rare mutations and high penetrance leads to the classic Mendelian pattern of dominant inheritance: if someone has a parent with HD, they inherit a 50% chance of developing HD themselves.

In fact these features appear so strongly in HD families that it is often taken as the 'prototype' of late-onset single-gene disorders. This is unfortunate, since there are many disorders of this type but few so extreme or clear-cut as HD. This apparent clarity means that HD has an extensive literature. It is also significant because although HD is caused by mutations at the same locus in a single gene, these mutations are variable in their effects, and the effect of that variability on the course of the disease is gradually becoming clearer. This is the first actuarial study of such a variable disorder.

The gene responsible, named the HD gene on chromosome 4, contains a repeated sequence of the bases cytosine, adenine and guanine (CAG for short) which encodes the amino acid glutamine in the gene product. Normally the CAG sequence is repeated 10–34 times, but it is unstable and tends to expand during meiosis, when the DNA is copied during the production of sperm and eggs. Once it reaches 40 or more repeats, HD results. Moreover, larger numbers of CAG repeats lead to lower ages at onset; the longest sequences observed (100 or more repeats) are associated with juvenile HD. Expansion is not inevitable (contraction is observed too) and it is more likely to occur in sperm than in eggs, meaning that HD may be more severe if inherited from the father. In Section 2 we describe the physiology of HD, and in Section 3 the mechanism of CAG repeats in the HD gene.

In Section 4, we propose models for the three features of HD epidemiology that most affect insurance costs, meaning either the premiums charged or the potential costs of adverse selection:

(a) We model the rate of onset as a function of CAG repeat length, based on the most comprehensive epidemiological study to date. The only existing actuarial model of HD (Smith (1998), described in Section 2.5) used aggregate rates of onset, ignoring CAG repeat length.
(b) We model post-onset survival, using a large study in which the definition of onset is consistent with the study used in (a) above.
(c) We model the distribution of CAG repeat lengths in the population, by applying a life table population model to estimates of the prevalence of CAG repeat lengths.

Our conclusions are in Section 5, and applications of the model are in Part II.

## 2. Huntington's Disease

### 2.1 *Mutation Frequency and Disease Prevalence*

The prevalence of HD is the proportion of people with symptoms of HD at any fixed calendar time. In Western European populations it is about 3–5 per 100,000 (Harper, 1996). Since others, especially at younger ages, will carry a HD mutation but not yet show symptoms, the frequency of the mutation in the population is much higher; Harper (1996) suggests about 2.5 times higher. Harper, Lim & Craufurd (2000) estimated the mutation frequency in the U.K. to be about 18.75 per 100,000, based on a disease prevalence of about 7.5 per 100,000.

### 2.2 *Physiology*

At a neurological level, HD is associated with degeneration in parts of the brain involved in the control of motor and mental functions. The degree of degeneration is linked to the severity of symptoms, which are both physical and psychological. Usually, slightly impaired muscle coordination is followed by progressively worse memory loss, disorganisation and personality changes. In detail, these can include:

(a) uncontrolled jerking movements, both voluntary and involuntary;
(b) slurred speech;
(c) depression, paranoid delusions and uncontrolled rage; and, in its later stages
(d) rigid joints and severe contractions leading to immobilisation and contortions,

Harper (1996) divided the progression of HD into three clinical stages, each corresponding roughly to a 5-year period of an overall 15-year course (Table 1). These could be relevant for certain insurances; broadly speaking, a claim under a CI or long-term care policy might be triggered between Stages 2 and 3, but a lot depends on individual underwriting practice.

The unified HD rating scale (UHDRS), composed of four components — motor performance, cognition, behaviour and functional capacity — has been introduced as a standardised measure of progression of the disease. Using this in a study of 960 patients, Marden *et al.* (2000) reported that age at onset, sex, weight and education had no effect on the rate of progression; depression was the only factor associated with more rapid decline. Similarly Feigin *et al.* (1995) found no correlation between rate of functional decline and age at onset of HD, body weight, gender of affected parent or history of neuroleptic use. These results suggest that the number of easily observable risk factors affecting progression is limited.

### 2.3 *Age at Onset of Huntington's Disease*

For actuarial applications, we need estimates of the rate of onset of HD at all ages of relevance to insurance, say 20–60 at least.

A basic problem is that, in common with many neurological disorders, the definition of 'onset' of HD is imprecise. A common definition is the time of the first definite abnormality, whatever its nature, recorded by a reliable witness (Harper, 1996). For our purposes, we need to be particularly careful:

Table 1: Stages of progression of Huntington's disease for a typical patient. Source: Harper (1996).

| Stage | Clinical Features |
|---|---|
| 1 | Presentation with initial neurological or psychiatric symptoms |
| | Main features remain similar to those at presentation |
| | Chorea more prominent than other motor abnormalities |
| | Patient largely independent for most activities |
| | Burden on family mainly result of psychiatric problems |
| | Death rare except for suicide |
| 2 | Motor disorders more generalised |
| | Physical disability becomes major |
| | Patient dependent on others for many activities |
| | Burden on family both physical and psychological |
| | Death often from unrelated causes |
| 3 | Severe generalised motor disorder |
| | Physical disability severe to total |
| | Patient completely dependent for all aspects of life |
| | Burden on carers mainly physical |
| | Death frequent at any point |

(a) when matching studies of onset with studies of post-onset mortality, if we must use studies based upon different populations; and
(b) when considering the timing of a CI insurance claim in relation to onset.

A summary of the larger studies (100 or more subjects) reporting mean or median age at onset is given in Table 2. These have no regard to the variation in HD associated with the variable nature of the mutation (many of them predate the sequencing of the HD gene) which must await Section 3.2. Those that report a range of ages at onset often show that juvenile HD has been included, which is a further reason for caution in using such studies; we are interested in the age at onset of HD conditional upon survival free of HD to adult ages, for which unconditional means could be misleading.

If a Normal distribution could be assumed then those studies which also estimate standard deviations could be useable. Several large studies have found that the distribution of ages at onset is very close to a Normal bell-curve (Bell, 1934; Wendt *et al.*, 1959; Wendt & Drohm, 1972) see Harper (1996). This feature was exploited in the model of Smith (1998), based on a similar observation by Roos *et al.* (1991), (see Section 2.5).

However, these studies must be interpreted with caution, both because of the variable definition of onset, and because genetic data present some unusual problems of analysis that have not always been allowed for.

(a) Rates of onset at older ages may be understated, and mean ages at onset biased towards lower ages, because of the exclusion of lives who are asymptomatic at the time of the study, but who later develop HD. Harper (1996) mentions two ways to correct for any bias:

Table 2: Mean age at onset of Huntington's disease in large studies (at least 100 subjects).

| Reference | Sample Size | Mean Age at Onset | Standard Deviation of Age at Onset | Range |
|---|---|---|---|---|
| Adams *et al.* (1988) | 611 | 38.66 | 11.69 | |
| Andrew *et al.* (1993) | 360 | 41.5 | 12.4 | 5–85 |
| Bell (1934) | 460 | 35.51 | 12.38 | 0–74 |
| Bolt (1970) | 265 | 42.7 | 13.2 | 5–72 |
| Brackenridge (1971) | 344 | 33.8 | | |
| Brinkman *et al.* (1997) | 728 | 41.0 | 13.0 | 4–84 |
| Brothers (1964) | 206 | 37.2 | | |
| Dewhurst *et al.* (1970) | 102 | 39.0 | | |
| Farrer & Conneally (1985) | 569 | 38.0 | 11.1 | 0–75 |
| Feigin *et al.* (1995) | 129 | 36.7 | 12.9 | 7–70 |
| Folstein *et al.* (1987) | 217 | 40.25 | 12.9 | 3–77* |
| Foroud *et al.* (1999) | 2,068 | 40.0 | 12.0 | 2–75 |
| Marden *et al.* (2000) | 960 | 40.8 | 12.4 | |
| Morrison *et al.* (1995) | 143 | 43.6 | 13.5 | 3–72 |
| Myers *et al.* (1985) | 243 | 40.9 | | 4–75 |
| Panse (1942) | 446 | 36.2 | 12.3 | |
| Reed *et al.* (1958) | 262 | 35.3 | 9.8 | |
| Roos *et al.* (1991) | 1,020 | 39.5 | 12.1 | |
| Stevens (1976) | 162 | 43.43 | 10.26 | 4–72 |
| Venters (1971) | 123 | 38.8 | 10.11 | 15–74 |
| Walker *et al.* (1981) | 204 | 41.2 | 12.7 | |
| Wallace (1972) | 144 | 39.5 | | |
| Wendt *et al.* (1959) | 762 | 43.97 | 10.9 | |
| Wendt & Drohm (1972) | 802 | 43.39 | 10.08 | |

* From Folstein (1989).

Table 3: Percentiles of survival times (years) since onset of HD based on the HD Roster in the U.S.A.. Sources: Foroud *et al.* (1999) and Roos *et al.* (1993).

| | | Foroud *et al.* (1999) | | | | Roos *et al.* (1993) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Percentiles of | | | | Percentiles of | | |
| | | No. of | Survival Times | | | No. of | Survival Times | | |
| Factor | Level | Subjects | 75% | 50% | 25% | Subjects | 80% | 50% | 20% |
| All | | 2,068 | 15.0 | 21.4 | 30.1 | 1,106 | 11.0 | 16.2 | 23.2 |
| Age at Onset | $< 20$ | 94 | 11.3 | 20.0 | 26.5 | 65 | 10.5 | 17.1 | 27.1 |
| | 20–34 | 537 | 14.8 | 21.3 | 29.7 | 274 | 12.3 | 17.0 | 25.0 |
| | 35–49 | 953 | 15.8 | 22.1 | 32.1 | 510 | 10.9 | 16.3 | 22.4 |
| | $\geq 50$ | 484 | 14.5 | 20.1 | 28.0 | 257 | 10.3 | 15.6 | 21.5 |
| Sex | Female | 1,074 | 15.3 | 22.0 | 31.8 | 532 | 11.2 | 17.1 | 23.3 |
| | Male | 994 | 14.8 | 20.8 | 29.0 | 574 | 10.5 | 15.5 | 22.5 |

(1) For data including asymptomatic at-risk individuals (in other words, censored data), classical life-table analysis may be used (Newcombe, 1981; Harper & Newcombe, 1992). This can substantially increase the probability of being a mutation carrier at older ages, so uncorrected age-at-onset curves are not recommended for use in genetic counselling.

(2) The other approach is to use older family histories, so that there are few censored observations. For example Adams *et al.* (1988) reported a mean age at onset of 38.66 years (Table 2) in respect of everyone in the study, but among lives born before 1921 the mean age at onset was 43.7 years: Wendt *et al.* (1959) and Wendt & Drohm (1972) estimated a mean age 43.4 years for onset in patients born between 1870 and 1900. On the other hand, such older information may be incomplete, less accurate and early-onset cases may be lost.

(b) Bias towards lower ages may also arise because of incomplete ascertainment, meaning that families with large numbers of affected persons, or unusually early onset, might be more likely to be included in studies. Very few studies are able to claim complete ascertainment (Morrison *et al.* (1995) is an example).

2.4 *Survival After Onset of Huntington's Disease*

In addition to the studies cited above, several others analyse survival times after onset of HD, and age at death. Harper (1996) cites 14 studies with mean survival times ranging from 10.6 to 17.1 years, and states that there is no clear difference by sex, age at onset or therapy.

Foroud *et al.* (1999) gave quartiles of survival times based on the very large HD Roster in the U.S.A. (Table 3). The differences between age at onset were marginally significant ($p = 0.025$) but small; those between males and females were not significant. These medians were much greater than the means previously reported, and also the medians given by Roos *et al.* (1993) (Table 3). This is almost certainly because the two studies used different definitions of onset. Roos *et al.* (1993) used onset of chorea, while Foroud

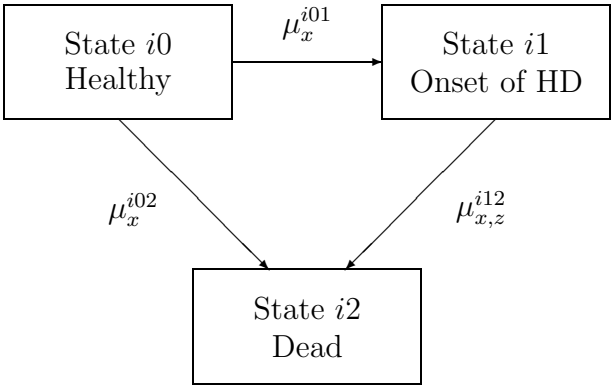


Figure 1: A multiple state model for Huntington's disease in life insurance, for a person with genotype $g_i$. The intensity $\mu_{x,z}^{i12}$ may depend on age $x$ and duration since onset $z$.

*et al.* onset of the earliest symptoms recalled once a diagnosis had been made. Roos *et al.* (1993) found that age at onset was not significant ($p = 0.013$) but sex might be, marginally ($p = 0.07$ or 0.008 depending on the test used).

There are no complications specific to HD but the combination of immobility, weight loss, tendency of aspirate food and general debility leaves patients generally vulnerable to disease. Pneumonia and cardiovascular diseases are among the commonest causes of death. However HD itself is generally the main cause of death, whatever appears on the death certificate.

2.5 *Smith's Model (1998): A Discrete-Time Semi-Markov Model*

Smith (1998) developed a discrete-time version of the model shown in Figure 1 and used it to price life insurance, both for individuals known to carry the mutation, and for those at risk of carrying it because of their family history. We describe it here because it was based on aggregate rates of onset of HD, ignoring the variations to be described in Section 3.

The principal genetical assumptions were that:

(a) the mutation is homogeneous in its effects (that is, it is simply either present or absent);
(b) there are no sporadic cases; HD only arises in families in which it is established; and
(c) the mutation is 100% penetrant.

As we will see in Section 3, the first of these is not true; there is significant variability. It is, however, quite common to ignore such variability in genetical studies, even in genetic epidemiology. The other two assumptions seem reasonable given the purpose of the model.

The model was fitted to results reported in Roos *et al.* (1991) and Roos *et al.* (1993), which predate the discovery of the variable CAG repeats. Data came from the Leiden register, which on 1 July 1990 included 1,106 lives with known age at onset. The data provided four 'pedigrees', namely the sex of the individual and the sex of their affected

parent.

Onset of HD in respect of known mutation carriers was modelled as a function of age, while the time to death after onset of HD was modelled as a function of duration since onset: both these features are consistent with many studies. For each 'pedigree', cumulative Normal distributions were fitted to the penetrance curves (the proportion of mutation carriers with HD by age $x$) and post-onset mortality curves (the proportion of affected lives dead $y$ years after onset); quite strong evidence exists for using the Normal in the former case (for example, Newcombe (1981)). One feature was that post-onset mortality depended on duration since onset alone, and not at all on age. Wilkie (2000) pointed out that this could cause the mortality rate to fall following onset at higher ages.

## 3. The Mechanism of HD: CAG Repeat Length in the HD Gene

### 3.1 *The Search for the Huntingtin Gene and the Development of Genetic Tests*

The search for the HD gene began in 1982. At that time, very little of the genome was sequenced; but the locations of a small number of identifiable DNA sequences called *markers*, scattered throughout the genome, were known. These markers were *polymorphic*, meaning that they varied between individuals. At meiosis, maternal and paternal chromosomes exchange lengthy segments of their DNA (this shuffling is why genes and not chromosomes are the basic unit of inheritance). Because the number of such segments is very small compared with the number of DNA bases in a chromosome, regions of DNA that are close together on a chromosome will remain together in the resulting sperm or egg, unless by chance one of the breaks between segments separates them. Therefore by finding one or more markers in an affected family, that were nearly always inherited along with HD, the HD gene could be pinned down to the region known to contain the marker. Almost at once, quite by chance, such a marker was found on the short arm of chromosome 4 (Gusella *et al.* 1983).

Markers provided a form of pre-symptomatic genetic testing within families; if a child carried the same variant of the marker as an affected forbear, it was possible, but not certain, that they carried the mutation too. Perhaps worse, not inheriting the same variant of the marker did not eliminate the risk of having the mutation. These tests gave results only in terms of probabilities; a refinement of Mendel's laws but not certainty. Take-up was relatively low (Harper, Lim & Craufurd, 2000). The most recent edition of Brackenridge & Elder (1998) — the authority on medical underwriting — treats this form of genetic test.

Finding the gene itself took another ten years and an enormous international collaboration (see Huntington's Disease Collaborative Research Group (1993) or Harper (1996, Chapter 8) for an account). Finally, it was the isolation of the gene that causes myotonic dystrophy (MD) that led to the breakthrough. MD and HD have some features in common, including anticipation, parental origin and unusual childhood forms (see Section 3.5); the cause of MD was found to be expanded trinucleotide repeats. The genetic code is written in triples of the DNA bases denoted A, C, G and T, called *trinucleotides*. Each triple (AAA, ACA, and so on) codes for an amino acid in a protein, or is a 'stop' instruction, so their linear sequence gives the linear sequence of amino acids in the gene product.

Table 4: CAG-repeat length categories and predicted phenotypes. Source: American College of Medical Genetics *et al.* (1998).

| Category | CAG-Repeat Range | Predicted Phenotype |
|---|---|---|
| Normal allele | ≤26 | Normal |
| Mutable normal allele | 27–35 | Normal |
| HD allele with reduced penetrance | 36–39 | Normal/HD |
| HD allele | ≥40 | HD |

Sometimes a long chain of the same trinucleotide occurs in a gene; in the MD gene it is CTG CTG CTG CTG .... Such chains may be unstable during meiosis, and the number of repeats liable to change between generations. If the number of CTG repeats in the MD gene exceeds about 2,000, the disease results (Pasternak, 1999).

When the region containing the HD gene was searched for similar structures, one was found in a gene already sequenced, called IT15: in this gene the trinucleotide sequence CAG, coding for the amino acid glutamine, was normally repeated 10–34 times, but was repeated 40 or more times in HD sufferers.

With this result, a genetic test was available of such reliability that it revolutionised the approach to HD in genetics clinics (Harper, 1996) and the uptake of testing rose substantially (Harper, Lim & Craufurd, 2000). Nevertheless, the devastating nature of HD and the lack of any treatment means that relatively few persons at risk opt for testing: Meiser & Dunn (2000) estimated 10–20%, while Harper, Lim & Craufurd (2000) estimated that by 1998 in the U.K., 18% of persons at 50% risk had been tested.

The American College of Medical Genetics *et al.* (1998) published a laboratory guideline for HD genetic testing, including the classification in Table 4.

### 3.2 *Age at Onset and CAG Repeat Length*

There is a strong correlation between age at onset and the number of CAG repeats, although these are not the only factors affecting age at onset (Andrew *et al.*, (1993); Ashisawa *et al.*, (1994); Brandt *et al.*, (1996); Brinkman *et al.*, (1997); Bruland *et al.*, (1999); Duyao *et al.*, (1993); Gusella & MacDonald, (1994); Kieburtz *et al.*, (1994); Luccotte *et al.*, (1995); Stine *et al.*, (1993); Trottier *et al.*, (1994)). The expanded CAG repeats is the unique cause of HD in all populations studied.

Brinkman *et al.* (1997) studied a cohort of 1,049 affected and asymptomatic at-risk persons, from many different countries. The number of CAG repeats was established for each person. For each number in the range 39–50, the cumulative probability of surviving without HD, at roughly quinquennial ages, was estimated by Kaplan-Meier methods, and 95% confidence intervals were also given. 'Age at onset' was defined as "... the first time a patient has either neurological or psychiatric symptoms that represented a permanent change from the normal state." These data for 40–50 CAG repeats are shown in Figures 2 and 3 in the form of penetrance estimates at selected ages:

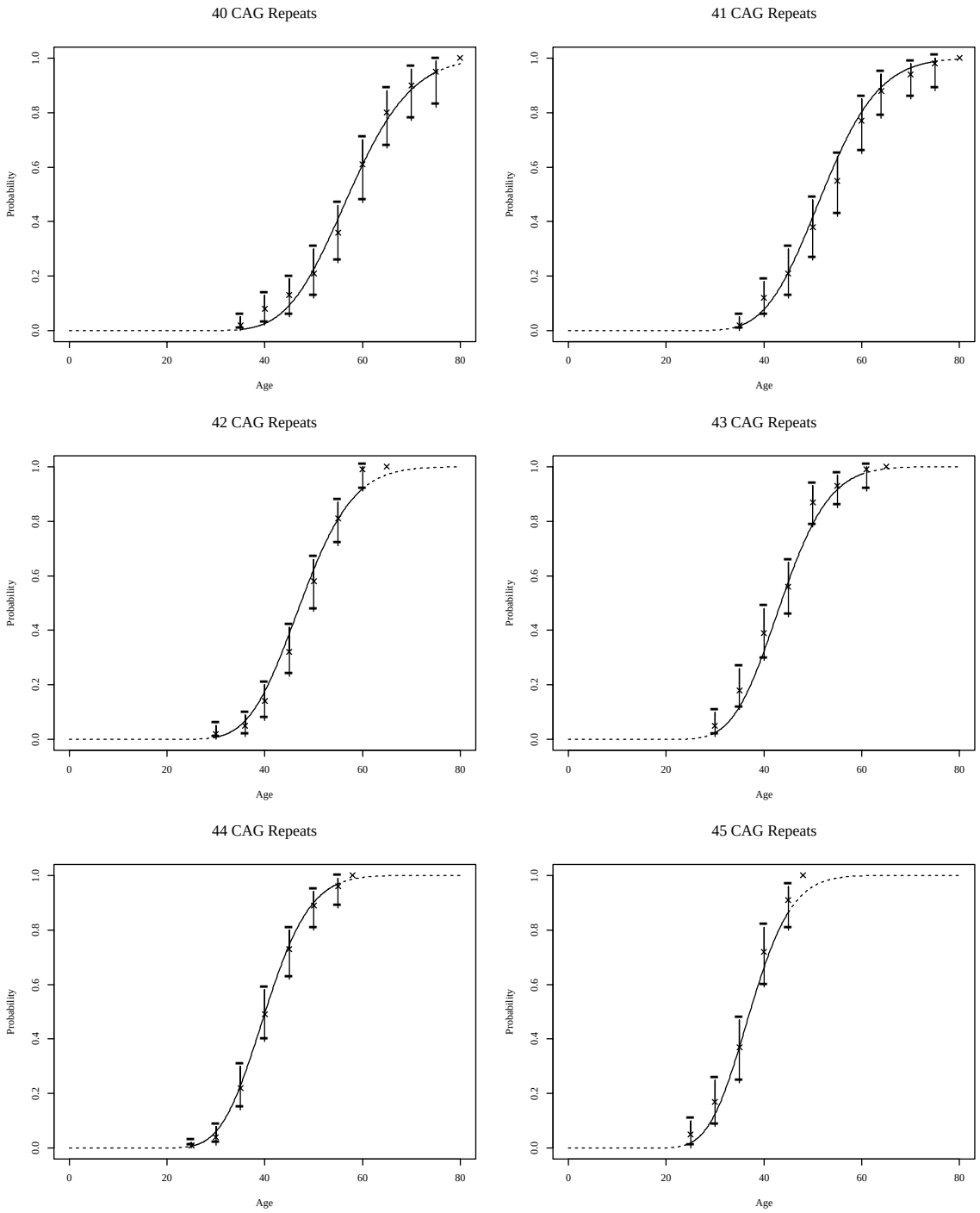


Figure 2: Penetrance estimates of onset of HD with 40–45 CAG repeats (crosses) and 95% confidence intervals, from Brinkman *et al.* (1997). Also shown are the fitted penetrance curves from Section 4.2.

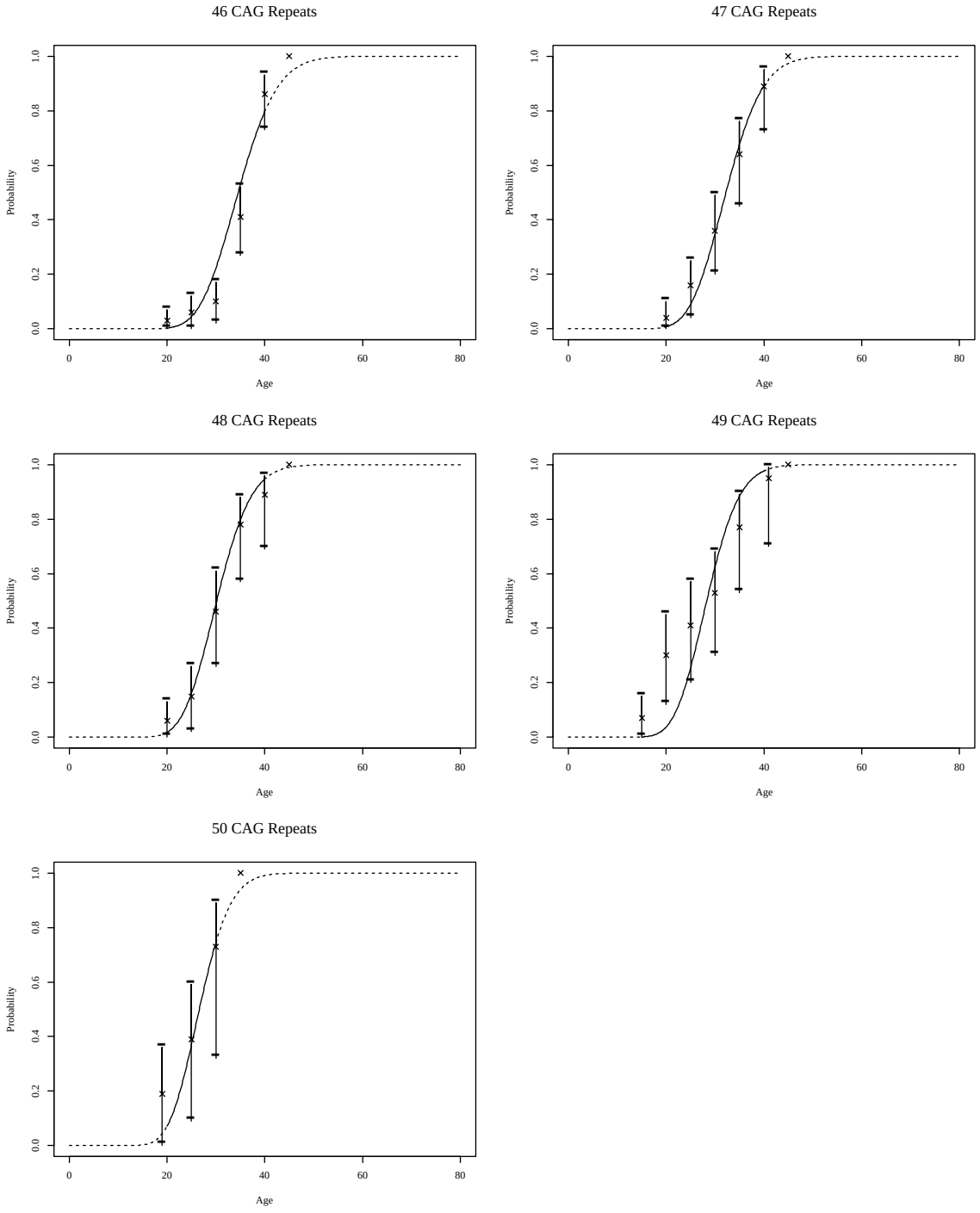


Figure 3: Penetrance estimates of onset of HD with 46–50 CAG repeats (crosses) and 95% confidence intervals, from Brinkman *et al.* (1997). Also shown are the fitted penetrance curves from Section 4.2.

Table 5: Frequency of CAG repeat sizes on normal and HD chromosomes. Source: Barron *et al.* (1993).

| Normal Chromosomes | | HD Chromosomes | |
|---|---|---|---|
| No. of CAG repeats | Frequency | No. of CAG repeats | Frequency |
| 8–13 | 0.033 | 35–39 | 0.113 |
| 14–18 | 0.709 | 40–44 | 0.631 |
| 19–23 | 0.189 | 45–49 | 0.187 |
| 24–28 | 0.054 | 50–54 | 0.045 |
| 29–33 | 0.015 | 55–59 | 0.018 |
| | | 60–64 | 0.006 |

$$\begin{aligned} \text{Penetrance at age } x &= \text{ Probability of having HD by age } x \\ &= 1 - \text{ Probability of surviving free of HD to age } x. \end{aligned}$$

(Also shown are fitted curves that will be described in Section 4.2.) In some cases, penetrance was 100% (everyone was observed to have HD by some finite age) and there are no confidence intervals for these point estimates. The main features were as follows:

(a) The survival functions were significantly different for each CAG repeat size; even a difference of a single CAG repeat was significant.
(b) No affected individuals had fewer than 36 CAG repeats.
(c) No individuals with more than 41 CAG repeats remained asymptomatic by age 65.
(d) Several individuals with 36–41 CAG repeats had no symptoms of HD within a normal expected life span; in other words, penetrance increased from 0% to 100% over this range.
(e) Two roughly linear relationships seemed to be present, one between log(age at onset) and log(number of CAG repeats); and one between (number of CAG repeats) and (median age at onset).

3.3 *Survival After Onset and CAG Repeat Length*

There is no clear correlation between the number of CAG repeats and rate of disease progression. Brandt *et al.* (1996) found no significant difference between survival times with fewer than 47 or more than 46 CAG repeats (67 subjects), though with unusually low disease durations. Foroud *et al.* (1999) reported shorter survival times with paternal rather than maternal transmission. The difference was statistically significant ($p = 0.66$) but very small.

3.4 *The Distribution of CAG Repeat Lengths*

A number of studies have reported the prevalence of CAG repeat lengths in sample populations. We show one example, which is typical. Barron *et al.* (1993) reported the frequency of CAG repeat sizes among 337 HD patients and 229 normal controls

Table 6: Normal distributions suggested for the age at onset of Huntington's disease.

| Reference | Population | Mean | Standard Deviation |
|---|---|---|---|
| Bell (1934) | Males | 36.05±0.51 | 12.16±0.36 |
| | Females | 35.17±0.60 | 12.51±0.42 |
| | All | 35.51±0.39 | 12.38±0.28 |
| Roos *et al.* (1991) | All | 39.5 | 12.1 |
| | Born before 1925 | 43.1 | 11.4 |
| Wendt & Drohm (1972) | All | 43.39 | 10.08 |
| Wilkie (2000) | Born before 1925 | 43.26* | 10.74* |

* Based on a subset of data presented in Roos *et al.* (1991).

from the Scottish population (Table 5). They observed that the peaks of the normal and HD ranges are 16 and 41 repeats respectively and there is no overlap of the two distributions. However, since this was a case-control study we would expect the numbers with a large number of CAG repeats to be depleted by deaths, so this might not estimate the distribution of CAG repeat sizes at birth.

Zühlke *et al.* (1993) reported the frequency of HD chromosomes by sex, finding no great differences. They found some cases of confirmed HD with fewer than 25 CAG repeats. A few other studies have reported confirmed cases of HD with fewer than 35 CAG repeats (Craufurd & Dodge, 1993; Kremer *et al.*, 1994; Snell *et al.*, 1993). The suggested explanation is that it might be evidence of a new mutation in the HD gene, or another gene involved in the huntingtin pathway.

Note that these studies might have missed asymptomatic individuals in the general population who have no affected relatives, but who have have a number CAG repeats in the range seen in affected persons with HD.

### 3.5 *Anticipation and Parental Origin*

In affected families, the age at onset of HD sometimes appears to fall in successive generations. This phenomenon is called 'anticipation'. Until quite recently, there was some doubt that this was a genuine feature of the disease, but the discovery of the expanding trinucleotide repeats revealed the mechanism at work. If the number of repeats is more likely to expand than to contract, and if a larger number of repeats leads to earlier onset, then anticipation is inevitable. Several studies (for example Ranen *et al.* (1995), Trottier *et al.* (1994)), report an increase in the number of CAG repeats especially when there is paternal transmission. An increase in the number of CAG repeats is related to an earlier age at onset in offspring than in parents; Snell *et al.* (1993) reported a mean change in age at onset between parent and child of 9.11 years with paternal transmission and 2.75 with maternal transmission.

Expansion is more likely to happen during the production of sperm than of eggs, for reasons as yet unknown. This leads to 'parental origin': HD is likely to be more severe, and anticipation clearer, if it is inherited from the father.

## 4. Modelling the Epidemiology of Huntington's Disease

### 4.1 *Estimating the Aggregate Rate of Onset*

As mentioned in Section 2.3, several authors have noted that the distribution of ages at onset appears to be close to a Normal bell-curve. Table 6 shows the parameterisations suggested by authors who specifically note this distribution; many of the other studies listed in Table 2 also cite means and standard deviations but do not necessarily recommend a Normal distribution. A Normal distribution does not have a strictly positive range, but provided its mean is greater than about three standard deviations this is immaterial.

### 4.2 *Estimating the Rate of Onset Depending on CAG Repeat Length 40–50*

In this section, for brevity, we define $R =$ CAG repeat length.

Brinkman *et al.* (1997) reported penetrance estimates of HD, and 95% confidence intervals, at (mostly) quinquennial ages for $R$ in the range 39–50 (the estimates for $R =$ 40–50 were shown in Figures 2 and 3). Each of these individually is based upon a much smaller sample than if they were pooled to find aggregate penetrance, so there is some irregularity as $R$ changes. However, certain features are clear; as $R$ increases:

(a) the earliest age at onset falls;
(b) the range of ages from earliest onset to 100% penetrance shortens.

CAG repeat length 39 falls into the intermediate range (Table 4) in which penetrance increases from zero to 100%, so we consider that separately in Section 4.3. Here we fit penetrance functions to CAG repeat lengths 40–50.

Because of the irregularities, we did not try to fit one-dimensional functions of age to the individual penetrance curves, but chose to fit a two-dimensional function of age and $R$ to the whole experience, giving an element of smoothing in both dimensions. The features (a) and (b) above suggested that a two-parameter distribution, with mean and variance both decreasing functions of $R$, would be suitable. The obvious candidate, because of Section 4.1, was a Normal$(\mu, \sigma)$ distribution with:

$$\mu = a - bR \qquad \text{and} \qquad \sigma = c - dR \tag{1}$$

for some positive parameters $a, b, c$ and $d$ (or any similar parameterisation). However, this did not give good results; no symmetric distribution would fit these data well. The best results were obtained with a Gamma distribution (which also has the slight advantage of a strictly positive range):

$$\text{Penetrance at age } x = \frac{\theta^\alpha}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} \exp(-t\theta) dt \tag{2}$$

where $\alpha = 48.1685 - 0.376508R$, $\theta = 0.051744R - 1.49681$ and $x \geq 0$. Fitting was by least squares, using approximate weights derived from the confidence intervals. The results are shown in Figures 2 and 3. The curves are shown as solid lines over the age range for which data were available, and as dotted lines where they are extended to lower and higher ages.

(a) No function that fitted $R =$ 40–50 at all well also fitted $R = 39$, so we deal with that separately later. This may not be surprising, as that is entering the zone between zero penetrance and complete penetrance.

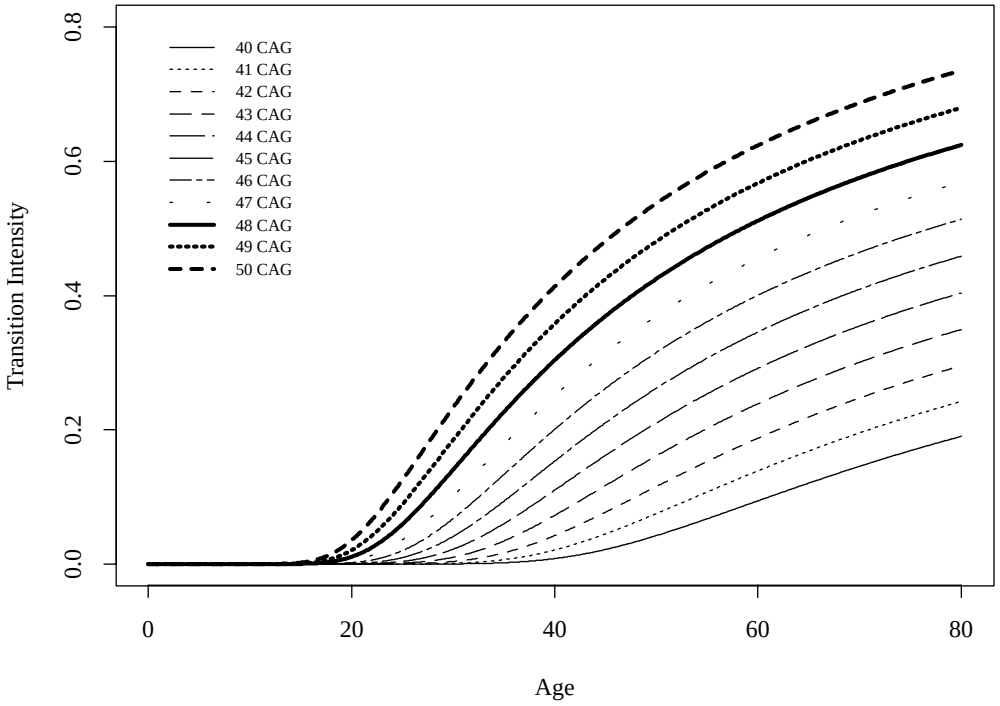


Figure 4: Rates (transition intensities) of onset of Huntington's disease depending on CAG repeat length, based on a Gamma model fitted to Brinkman *et al.* (1997).

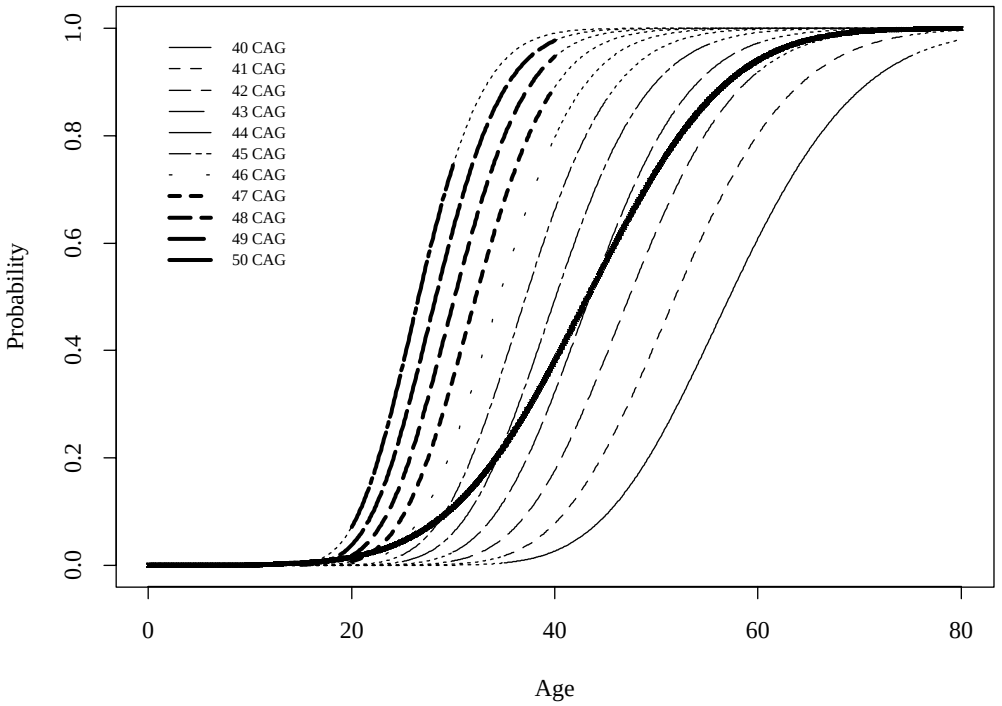


Figure 5: Penetrance of HD mutations depending on CAG repeat length, based on a Gamma model fitted to Brinkman *et al.* (1997), compared with the cumulative Normal model for the aggregate penetrance suggested by Wilkie (2000) (bold line).

(b) The fits are quite good except for $R = 46$ and 49, based on 63 and 30 affected subjects respectively. There is other evidence that $R = 49$ was out of line with the other observations; the minimum age at onset fell from 35 years with $R = 40$ to 16 years with $R = 50$ except for a jump down to 13 years with $R = 49$. For use in actuarial or demographic applications, therefore, the smoothed function based on all the data may be preferred to the individual experience with $R = 49$.

Figure 4 shows the intensities of onset for $R = 40$–50 based on Equation (2), and Figure 5 compares the fitted penetrance curves with the Normal model of aggregate penetrance from Wilkie (2000). This shows clearly how much heterogeneity is ignored by any model that uses aggregate rates of onset.

4.3 *Fewer Than 40 CAG Repeats*

Roughly, penetrance increases from zero with 35 CAG repeats to 100% with 40 CAG repeats (Table 4). Brinkman *et al.* (1997) estimated penetrance with 39 CAG repeats, but from a small sample. It will always be more difficult to obtain reliable estimates of penetrance in this intermediate range:

(a) Because of the mechanism of trinucleotide expansion, it will contain the first people in previously non-HD families to have more than 35 CAG repeats, and they are much more likely to be missed or misdiagnosed. Falush *et al.* (2000) found that

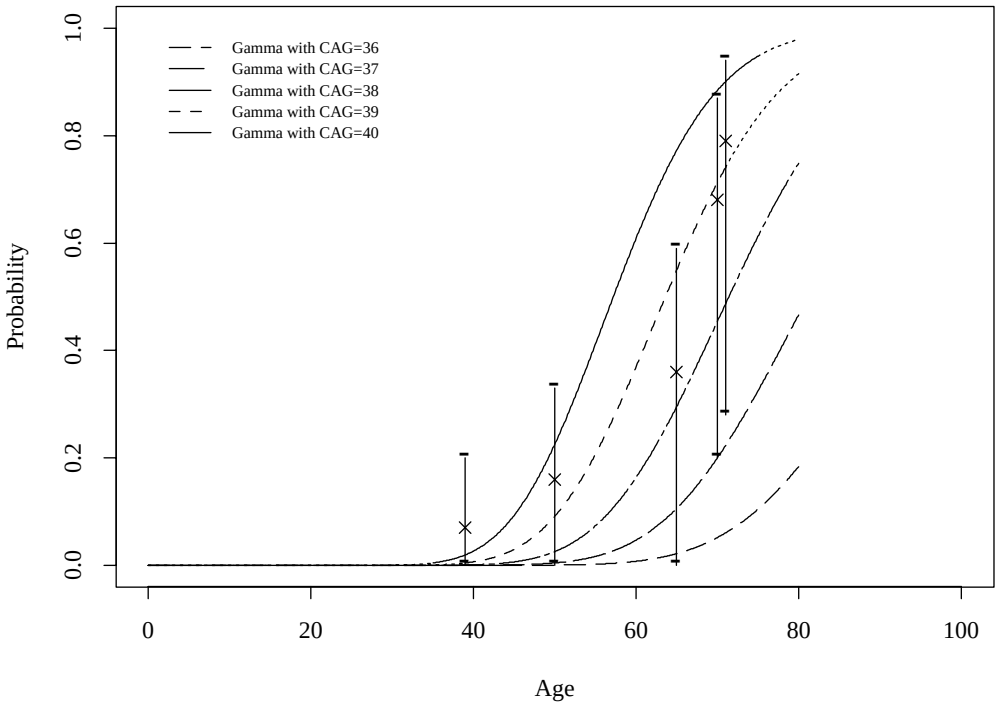


Figure 6: Estimated penetrance with 39 CAG repeats (Brinkman *et al.*, 1997) compared with the Gamma model fitted to 40–50 CAG repeats and extrapolated to 36–39 CAG repeats.

ascertainment fell to less than 50% with 40 repeats and less than 5% with 36–38 repeats.

(b) Incomplete penetrance means smaller numbers of cases, leading to less reliable estimates with any given sample size, and more danger of ascertainment bias.

Therefore, it is interesting to note the properties of the Gamma function fitted to 40–50 CAG repeats, if it is extrapolated to 36–39 CAG repeats. The resulting penetrance curves are shown in Figure 6, along with the estimate for 39 CAG repeats from Brinkman *et al.* (1997).

(a) Given the very wide confidence intervals, the Gamma function with $R = 39$ or 38 might be plausible.

(b) The Gamma function gives the right sort of overall penetrance by about age 60, (as high as we need for many insurance questions) being close to zero with $R = 36$. Beyond about age 70 there are no data.

Using this Gamma function in the range $R = 36$–39 would be speculative. It would be impossible, for example, to draw any conclusions that could be used in counselling. It might, nevertheless, give useful guidance on the general level of insurance risk in respect of persons with intermediate test results, as long as its limitations are borne in mind.

4.4 *Survival After Onset of Huntington's Disease*

Smith (1998) (see Section 2.5) fitted Normal distributions to percentiles of survival times given by Roos *et al.* (1993), which depended on parental origin. A major advantage of using these mortality rates in his study was that rates of onset used came from Roos *et al.* (1991), based upon the same data, ensuring that the definitions of age at onset were consistent (in this case, the age at which choreatic movements become manifest). Table 3 showed significant differences between the survival distributions in Roos *et al.* (1993) and Foroud *et al.* (1999), suggesting that definition of age at onset is extremely important. Note also that Roos *et al.* (1993) did not exclude cases of juvenile HD, which are:
(a) associated with exceptionally large CAG repeats, hence with paternal origin; and
(b) irrelevant for our purposes.

Given the definition of onset used by Brinkman *et al.* (1997) (see Section 3.2) it is doubtful if the post-onset rates of mortality based on Roos *et al.* (1993) would be consistent with the rates of onset we have used. As mentioned in Section 2.4, Foroud *et al.* (1999) defined age at onset as the age at which initial symptoms are detected, not necessarily chorea. This is clearly closer to the definition used in Brinkman *et al.* (1997).

Foroud *et al.* (1999) also estimated probabilities of survival after onset of HD as a function of duration since onset, for four age-at-onset groups (less than 20, 20–34, 35–49, 50 and over). Juvenile HD is therefore excluded by ignoring the first of these. The estimates are shown in Figure 7. That for age at onset 50 and over was slightly, though significantly, different from those for ages 20–34 and 35–49, but the authors noted that this could just be the usual age-related mortality difference.

For these reasons, we will base post-onset survival in our model on Foroud *et al.* (1999), using the following graduations (also shown in Figure 7) in which $S(d)$ is the probability of surviving for $d$ years since onset. For age at onset 20–34:

$$1 - S(d) = \frac{0.174219^{4.11789}}{\Gamma(4.11789)} \int_0^d t^{3.11789} e^{-0.174219t} dt. \tag{3}$$

For age at onset 35–49:

$$1 - S(d) = \frac{0.177225^{4.35046}}{\Gamma(4.35046)} \int_0^d t^{3.35046} e^{-0.177225t} dt. \tag{4}$$

For age at onset 50 and over:

$$1 - S(d) = \frac{0.183372^{4.1465}}{\Gamma(4.1465)} \int_0^d t^{3.1465} e^{-0.183372t} dt. \tag{5}$$

4.5 *The Distribution of CAG Repeat Lengths*

The potential for adverse selection in insurance related to HD depends on the proportion of the population at risk of HD as well as the risk of onset. Since the latter depends on CAG repeat length, we need the distribution of these in the population. However, most studies give mutation prevalences from cross-sectional surveys of affected or at-risk persons or families. For example, Table 5 gave mutation prevalences from Barron *et*

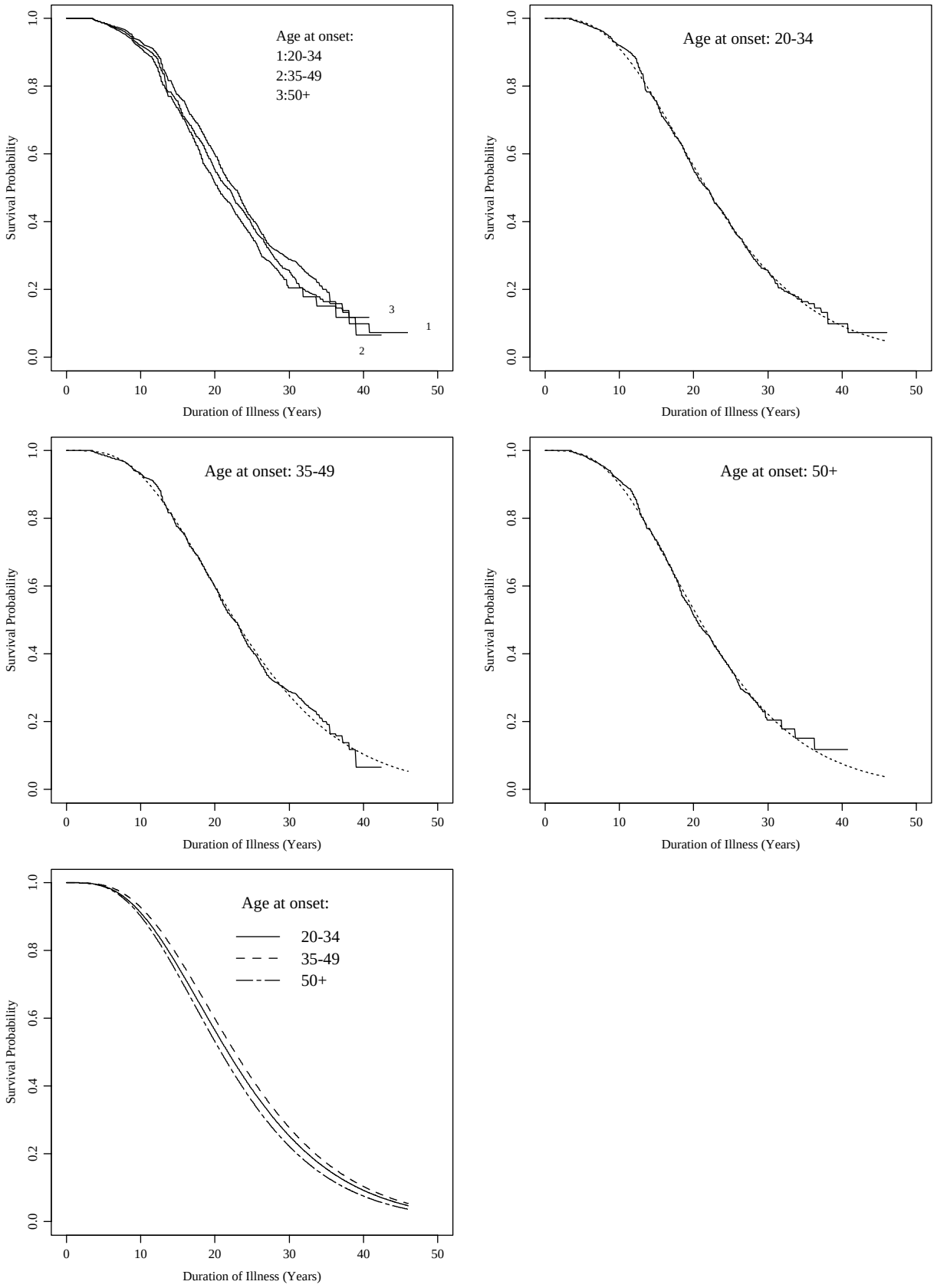


Figure 7: Probability of survival as a function of duration since onset of HD, based on Foroud *et al.* (1999). Shown are: Kaplan-Meier estimates for three age-at-onset groups (top left); graduations of each (Section 4.4); and a comparison of the three graduations.

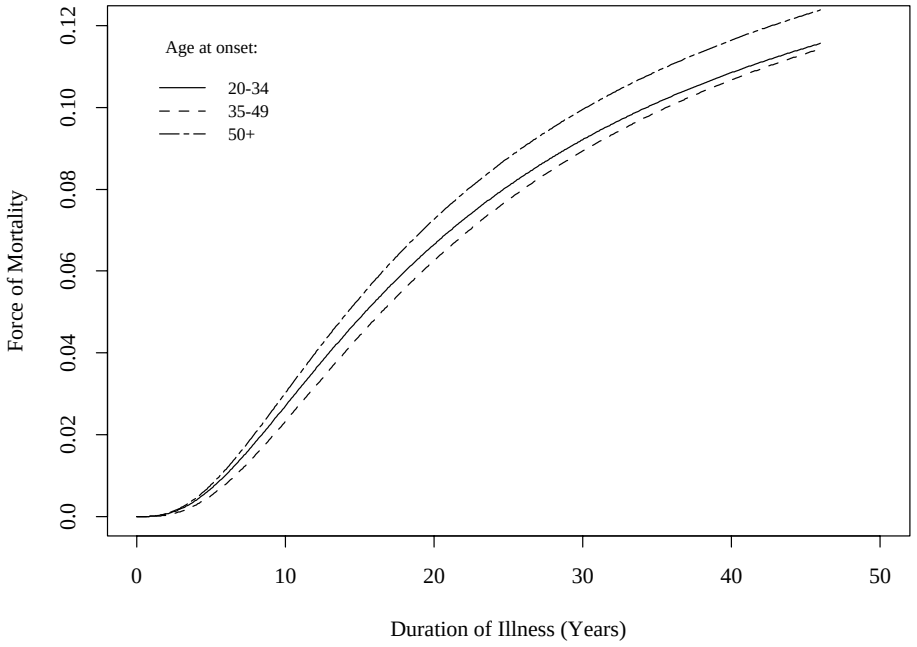


Figure 8: Force of mortality as a function of duration since onset of HD, based on Foroud *et al.* (1999).

*al.* (1993), for groups of CAG repeat lengths. This included symptomatic and asymptomatic tested individuals. We can assume that the groups with high numbers of CAG repeats have been more severely depleted, and that these were more common at birth. Given rates of onset representative of these groups, we can work back from prevalences to obtain approximate frequencies at birth.

(a) Suppose there are $M$ groups of CAG repeat lengths. For each one, consider the three-state model in Figure 1; let the occupancy probabilities be:

$$
\begin{aligned}
{}_tp_x^{ijk} &= \text{P[In state } ik \text{ at age } x+t \mid \text{In state } ij \text{ at age } x] \qquad (j,k \neq 1) \qquad (6)\\
{}_{t,z}p_x^{ij1} &= \text{P[In state } i1, \text{ duration } z \text{ at age } x+t \mid \text{In state } ij \neq i1 \text{ at age } x] \qquad (7)\\
{}_tp_{x,z}^{i1k} &= \text{P[In state } ik \text{ at age } x+t \mid \text{In state } i1, \text{ duration } z \text{ at age } x] \qquad (8)
\end{aligned}
$$

(b) In a stable population with birth rate 1 per annum, of which a proportion $p_i$ are in the $i^{th}$ group ($p_1 + \ldots + p_M = 1$) the expected number of symptomatic and asymptomatic persons in the $i^{th}$ group at any time is:

$$
p_i \int_0^\infty ({}_xp_0^{i00} + {}_xp_0^{i01})dx = p_i \int_0^\infty \left( {}_xp_0^{i00} + \int_0^x {}_yp_0^{i00}\mu_y^{i01}{}_{x-y}p_{y,0}^{i11}\, dy \right) dx. \qquad (9)
$$

We can compute these integrals, given rates of onset and of survival after onset. Clearly we can do the same for symptomatic and asymptomatic persons separately.

Table 7: Summary of the distributions of CAG repeat length in the 'affected' range. CAG repeats 36–50. Quantity actually fitted is (CAG − 35).

| Study | Persons | Mean | Variance | Variance/Mean |
|---|---|---|---|---|
| Brinkman *et al.* (Asymptomatic) | 230 | 6.89 | 9.14 | 1.33 |
| Brinkman *et al.* (Symptomatic) | 668 | 8.66 | 7.33 | 0.85 |
| Brinkman *et al.* (Combined) | 898 | 8.21 | 8.38 | 1.02 |
| Barron *et al.* (Combined) | 316 | 7.34 | 7.18 | 0.98 |
| Craufurd & Dodge (Symptomatic) | 193 | 9.93 | 7.45 | 0.75 |
| Kremer *et al.* (Symptomatic) | 878 | 9.55 | 7.27 | 0.76 |

(c) If we observe $P_i$ alleles (or chromosomes or symptomatic and/or asymptomatic persons) in the $i^{th}$ group, of unknown but presumed representative age distribution, we assume each $P_i$ is proportionate to the corresponding member of Equations (9), and solve for the ratios $p_2/p_1, \ldots, p_M/p_1$ along with $p_1 + \ldots + p_M = 1$. (In practice, we might choose the group with the largest $P_i$ as the denominator.)

We estimated $p_i$ in this way from four studies. Brinkman *et al.* (1997) report the numbers of CAG repeats among asymptomatic and symptomatic persons separately, but group 51 or more CAG repeats together. Barron *et al.* (1993) present the numbers of CAG repeats among asymptomatic and symptomatic persons combined, while Craufurd & Dodge (1993) and Kremer *et al.* (1994) studied symptomatic persons only. Brinkman *et al.* (1997) was the only study to give the actual numbers, the others published histograms from which reasonable estimates could be read. These show that the distribution of CAG repeat length is bimodal, with a clear peak in the 'normal' range and another just above 40 CAG repeats; we are only concerned with the latter. Table 7 summarises the distributions of (CAG repeat length − 35). It suggests that $p_i$ is underdispersed among symptomatic persons and overdispersed among asymptomatic persons. Possibly this reflects the reported underascertainment of CAG repeats in the range 36–39 (Falush *et al.*, 2000).

We chose to use Brinkman *et al.* (1997) combined (hence Equations (9)), despite their grouping 50 or more CAG repeats, because our rates of onset are based on these data, and we have no reason to prefer to use samples of only asymptomatic or only symptomatic persons. Figure 9 (a) shows the resulting values of $p_i$, and for comparison Figure 9 (c) shows those based on Barron *et al.* (1993). They are reasonably consistent with each other, and with the suggestion of underascertainment in the range 36–39. For our purposes, we need to smooth them, bearing in mind that our samples are truncated at 50 CAG repeats; we found the following translated-truncated Poisson distribution gave acceptable results, shown in Figure 9 (b) and (d):

$$\hat{p_i} = \frac{\exp(-\lambda)\frac{\lambda^{(i-30)}}{(i-30)!}}{\sum_{r=36}^{50}\exp(-\lambda)\frac{\lambda^{(r-30)}}{(r-30)!}} \tag{10}$$

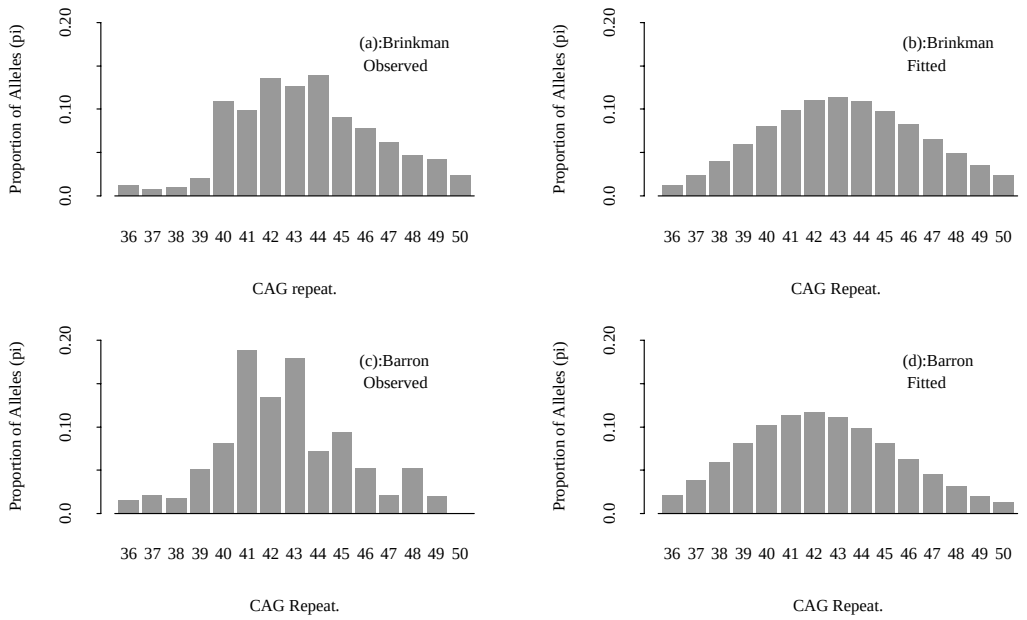


Figure 9: Estimated distribution of CAG repeat lengths at birth, based on numbers of asymptomatic and symptomatic individuals in Brinkman *et al.* (1997) and Barron *et al.* (1993). Fitted values are based on a truncated Poisson model.

with $\lambda = 14.0395$ for Brinkman *et al.* (1997) and $\lambda = 13.4451$ for Barron *et al.* (1993). Formal goodness-of-fit tests suggest a poor fit, mainly because of the results for 36–39 CAG repeats, but we accept the apparent discrepancy because of the suspected underascertainment. Any error in this part of the range has only a small effect on insurance costs. Table 8 shows the values of $p_i$ estimated as in Equation (9) from Brinkman *et al.* (1997), and the fitted values.

Brinkman *et al.* (1997) reported 65 individuals with 50–121 CAG repeats (out of 963 with 36 or more), but did not estimate rates of onset for them. Therefore, our model of onset and of the distribution of CAG repeats is necessarily limited to 36–50 CAG repeats, but we thereby exclude only a small proportion of the at-risk population, (among which are cases of juvenile onset that should be excluded for our purposes anyway) and penetrance is so high with 50 CAG repeats that we are not omitting any features of importance.

Given this assumed distribution of CAG repeat lengths, we can find the aggregate penetrance of HD mutations implied by the model. Figure 10 shows this aggregate penetrance based upon 40–50 CAG repeats only, and upon 36–50 CAG repeats. The latter is quite close to the Normal model suggested by Wilkie (2000), while the inclusion of intermediate alleles reduces the penetrance markedly at higher ages. This would be consistent with underascertainment of intermediate alleles in Roos *et al* (1991), used by Wilkie (2000), or it might indicate that we were wrong to accept the fitted model shown in Figure 9 (b) where we explicitly assumed that there was underascertainment, and removed it.

Table 8: Estimated distribution of CAG repeat lengths at birth, based on numbers of asymptomatic and symptomatic individuals in Brinkman *et al.* (1997).

| CAG repeats | Estimated | Fitted | CAG repeats | Estimated | Fitted |
|---|---|---|---|---|---|
| 36 | 0.0122 | 0.0124 | 44 | 0.1398 | 0.1094 |
| 37 | 0.0075 | 0.0238 | 45 | 0.0902 | 0.0980 |
| 38 | 0.0104 | 0.0400 | 46 | 0.0780 | 0.0824 |
| 39 | 0.0201 | 0.0598 | 47 | 0.0619 | 0.0652 |
| 40 | 0.1085 | 0.0804 | 48 | 0.0469 | 0.0487 |
| 41 | 0.0988 | 0.0983 | 49 | 0.0416 | 0.0344 |
| 42 | 0.1349 | 0.1101 | 50 | 0.0229 | 0.0232 |
| 43 | 0.1264 | 0.1139 | | | |

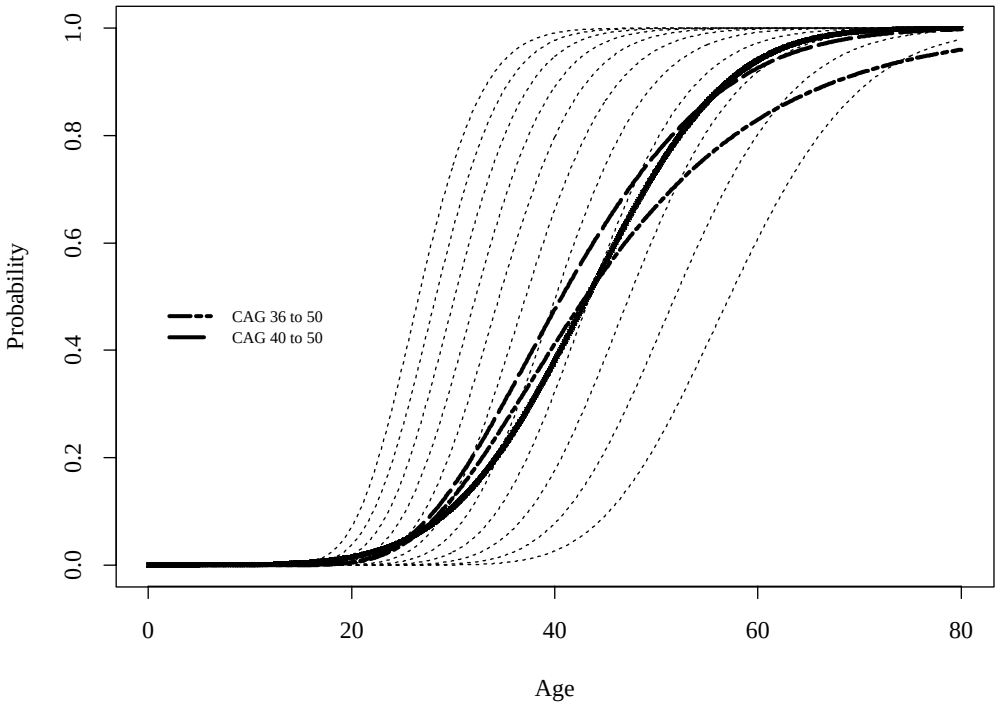


Figure 10: Aggregate penetrance of HD mutations implied by the model, based on 36–50 and 40–50 CAG repeats, compared with the cumulative Normal model for the aggregate penetrance suggested by Wilkie (2000) (solid line). Penetrances of 40–50 CAG repeats individually are shown as dotted lines.

## 5. Conclusions

Smith (1998) proposed a discrete-time semi-Markov model of HD onset and post-onset mortality, and applied it to life insurance pricing. He used aggregate rates of onset, not distinguishing CAG repeat length whose effect is now known to be considerable. Our aim was to model the effect of CAG repeat length, to study CI insurance as well as life insurance, and also to consider the possible costs of adverse selection arising from non-disclosure of genetic test results and/or family history.

Brinkman *et al.* (1997) provided Kaplan-Meier estimates of age at onset of HD for 39–50 CAG repeats, and we fitted a two-dimensional surface to these for 40–50 CAG repeats, the penetrance for each CAG repeat number being a Gamma function. Furthermore, this function behaved sensibly at the relevant ages when extrapolated to fewer CAG repeats, giving almost zero penetrance at age 60 with 36 CAG repeats.

Foroud *et al.* (1999) gave Kaplan-Meier estimates of duration-dependent post-onset mortality rates, for various ages at onset. There was no dependence on the CAG repeat length. Most important, the definition of onset appeared to be consistent with that used by Brinkman *et al.* (1997) (unlike other large studies of post-onset mortality) and we adopted these rates, after smoothing them.

For applications in which the available information indicates a risk of HD but not the CAG repeat length, the distribution of CAG repeat lengths in the population is needed. We estimated this by applying a stationary population model (based on the onset and mortality rates above) to the prevalences of CAG repeat lengths in Brinkman *et al.*'s (1997) sample, and smoothing the resulting proportions. The smoothing intentionally compensated for apparent underascertainment of fewer than 40 CAG repeats.

## Acknowledgements

This work was carried out at the Genetics and Insurance Research Centre at Heriot-Watt University, which receives funding from a consortium of insurance companies through the Association of British Insurers. We thank the sponsors for funding, and members of the Steering Committee for helpful comments at various stages. We are grateful to Professor Peter Harper for his help.

## References

Adams P., Falek, A. & Arnold, J. (1988). Huntington disease in Georgia: age at onset. *American Journal of Human Genetics*, **43**, 695–704.

American College of Medical Genetics/American Society of Human Genetics, Huntington Disease Genetic Testing Working Group (1998). Laboratory guidelines for Huntington's disease genetic testing. *American Journal of Human Genetics*, **62**, 1243–1247.

Andrew, S., Goldberg, P., Kremer, B., Telenius, H., Theilmann, J., Adam, S., Starr, E., Squitieri, F., Lin, B., Kalchman, M., Graham, R., & Hayden, M. (1993). The relation between trinucleotide (CAG) repeats length and clinical features of Huntington's disease. *American Journal of Human Genetics*, **43**, 695–704.

Ashizawa, M.D., Wong, L.-J.C., Richards, C.S., Caskey, C.T. & Jankovic, J. (1993). 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, D. (1993). A study of the Huntington's disease associated trinucleotide repeat in the Scottish population. *Journal of Medical Genetics*, **30**, 1003–1007.

Bell, J. (1934). Huntington's Chorea. *Treasury of Human Inheritance*, **IV**, 1–67.

Bolt, J.M. (1970). Huntington's chorea in the West of Scotland. *British Journal of Psychiatry*, **116**, 259–270.

Brackenridge, C. (1971). A genetic and statistical study of some sex-related factors in Huntington's disease. *Clinical Genetics*, **2**, 267–286.

Brackenridge, R. & Elder, J. (1998). *Medical selection of life risks. Fourth Edition.* Macmillan.

Brandt, J., Bylsma, F., Gross, R., Stine, O., Ranen, N. & Ross, C. (1996). Trinucleotide repeat length and clinical progression in Huntington's disease. *American Academy Neurology*, **46**, 527–531.

Brinkman, R., Mezei, M., Theilmann, J., Almqvist, E. & Hayden, M. (1997). The likelihood of being sffected with Huntington disease by a particular age, for a specific CAG size. *American Journal of Human Genetics*, **60**, 1202–1210.

Brothers, C.R.D. (1964). Huntington's chorea in Victoria and Tasmania. *Journal of the Neurological Sciences*, **1**, 405–420.

Craufurd, D. & Dodge, A. (1993). Mutation size and age at onset in Huntington's disease. *Journal of Medical Genetics*, **30**, 1008–1011.

Dewhurst, K., Oliver, J.E. & McKnight, L. (1970). Huntington's Disease. *British Journal of Psychiatry*, **116**, 255–258.

Duyao, M., Ambrose, C., Myers, R., Novelletto, A., Persichetti, F., Frontali, M., Folstein, S., Ross, C., Franz, M., Abbott, M., Gray, J., Conneally, P., Young, A., Penney, J., Hollingsworth, Z., Shoulson, I., Lazzarini, A., Falek, A., Koroshetz, W., Sax, D., Bird, E., Vonsattel, J., Bonilla, E., Alvir, J., Bickham, J., Cha, J., Dure, L, Gomez, F., Ramos, M., Sanchez-Ramos, J., Snodgrass, S., de Young, M., Wexler, N., Moscowitz, G., Penchaszadeh, G., MacFarlane, H., Anderson, M., Jenkins, B., Srinidhi, J., Barnes, G., Gusella, J. & Macdonald, M. (1993). Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genetics*, **4**, 387–392.

Falush, D., Almqvist, E.W., Brinkmann, R.R., Iwasa, Y. & Hayden, M.R. (2000). Measurement of mutational flow implies both a high new-mutation rate for Huntington disease and a substantial underascertainment of late-onset cases. *American Journal of Human Genetics*, **68**, 373–385.

Farrer, L. & Conneally, P. (1985). A genetic model for age at onset in Huntington disease. *American Journal of Human Genetics*, **37**, 350–357.

Feigin, A., Kieburtz, K., Bordwell, K., Como, P., Steinberg, K., Sotack, J., Zimmerman, C., Hickey, C., Ormen, C. & Shoulson, I. (1995). Functional decline in Huntington's disease. *Movement Disorders*, **10**, 211–214.

Folstein, S. (1989). *Huntington's disease. A disorder of families.* The Johns Hopkins University Press.

Folstein, S.E., Chase G.A., Wahl, W.E., McDonnell, A.M. & Folstein, M.F. (1987). Huntington disease in Maryland: Clinical aspects of racial variation. *American Journal of Human Genetics*, **41**, 168–179.

Foroud, T., Gray, J., Ivashina, J. & Conneally, M. (1999). Differences in duration of Huntington's disease based on age at onset. *Journal of Neurology, Neurosurgery and Psychiatry*, **66**, 52–56.

Gusella, J., Wexler, N., Conneally, M., Naylor, S., Anderson, M., Tanzi, R., Watkins, P., Ottina, K., Wallace, M., Sakaguchi, A., Young, A., Shoulson, I., Bonilla, E. & Martin, J. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*, **306**, 234–238.

Gusella, J.F. & MacDonald, M.E. (1994). Huntingtons's disease and repeating trinucleotides. *New England Journal of Medicine*, **330**, 1450–1451.

Harper, P.S. (1996). *Huntington's disease*. W.B. Saunders.

Harper, P.S. & Newcombe, R.G. (1992). Age at onset and life table risks in genetic counselling for Huntington's disease. *Journal of Medical Genetics*, **29**, 239–242.

Harper, P.S., Lim, C. & Craufurd, D. (2000). Ten years of presymptomatic testing for Huntington's disease: the experience of the UK Huntington's Disease Prediction Consortium. *Journal of Medical Genetics*, **37**, 657–571.

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.

Kieburtz, K., MacDonald, M., Shih, C., Feigin, A., Steinberg, K., Bordwell, K., Zimmerman, C., Srinidhi, J., Sotack, J., Gusella, J. & Shoulson, I. (1994). Trinucleotide repeat length and progression of illness in Huntington's disease. *Journal of Medical Genetics*, **31**, 872–874.

Kremer, B., Goldberg, P., Andrew, S., Theilmann, J., Telenius, H., Zeisler, J., Squitieri, F., Lin, B., Bassett, A., Almqvist, E., Bird, T. & Hayden, M. (1994). A Worldwide study of the Huntington's disease mutation. The sensitivity and specificity of measuring CAG repeats. *New England Journal of Medicine*, **330**, 1401–1406.

Lucotte, G., Turpin, J.C., Riess, O., Epplen, J.T., Siedlaczk, I., Loirat, F. & Hazout, S. (1995). Confidence intervals for predicted age of onset, given the size of $(CAG)_n$ repeat, in Huntington's disease. *Human Genetics*, **95**, 231–232.

Macdonald, A.S. (1999). Modeling the impact of genetics on insurance. *North American Actuarial Journal*, **3:1**, 83–101.

Macdonald, A.S. (2003). Genetics and insurance: what have we learned so far?. *To appear in Scandinavian Actuarial Journal*.

Marden, K., Zhao, H., Myers, R., Cudkowicz, M., Kayson, E., Kieburtz, K., Orme, C., Paulsen, J., Penney, J., Siemers, E., Shoulson, I. & The Huntington Study Group (2000). Rate of functional decline in Huntington's disease. *Neurology*, **54**, 452–458.

Meiser, B. & Dunn, S. (2000). Psychological impact of genetic testing for Huntington's disease: an update of the literature. *J. Neurol. Neurosurg. Psychiatry*, **69**, 574–578.

Morrison, P., Johnston, W. & Nevin, N. (1995). The epidemiology of Huntington's disease in Northern Ireland. *Journal Medical Genetics*, **32**, 524–530.

Myers, R., Sax, D.S., Schoenfeld, M., Bird, E., Wolf, P., Vonsattel, J.P., White, R.F. & Martin, J.B. (1985). Late onset of Huntington's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, **48**, 530–534.

Newcombe, R.G. (1981). A life table for onset of Huntington's Chorea. *Annals of Human Genetics*, **45**, 375–385.

Panse, F. (1942). *Die erbchorea; eine klinish-genetisch studie.* Leipzing:Thieme.

Pasternak, J.J. (1999). *An introduction to human molecular genetics.* Fitzgerald Science Press, Bethesda, Maryland.

Ranen, N., Stine, C., Abbott, M., Sherr, M., Codori, A., Franz, M., Chao, N., Chung, A., Pleasant, N., Callahan, C., Kasch, L., Ghaffari, M., Chase, G., Kazazian, H., Brandt, J., Folstein, S. & Ross, C. (1995). Anticipation and instability of IT-15 $(CAG)_N$ repeats in parent-offspring pairs with Huntington disease. *American Journal of Human Genetics*, **57**, 593–602.

Reed, T.E., Chandler, J.H., Hughes, E.M. & Davidson, R.T. (1958). Huntington's chorea in Michigan. *American Journal of Human Genetics*, **10**, 201–225.

Roos, R.A.C., der Vlis, M.V., Hermans, J., Elshove, H.M., Moll, A.C., van de Kamp, J.J.P. & Bruyn, G.W. (1991). Age at onset in Huntington's disease: effect of line of inheritance and patient's sex. *Journal of Medical Genetics*, **28**, 515–519.

Roos, R., Hermans, J., Vegter-van der Vlis, M., van Ommen, G. & Bruyn, G. (1993). Duration of illness in Huntington's disease is not related to age at onset. *Journal of Neurology, Neurosurgery and Psychiatry*, **56**, 98–100.

Smith, C. (1998). *Huntington's chorea: A mathematical model for life insurance.* Unpublished manuscript, Swiss Re, Zurich.

Snell, R., MacMillan, J., Cheadle, J., Fenton, I., Lazarou, L., Davies P., MacDonald, M., Gusella, J., Harper, P. & Shaw, D. (1993). Relationship between trinucleotide repeat expansion and phenotypic variaton in Huntington's disease. *Nature Genetics*, **4**, 393–397.

Stevens, D. (1976). *Huntington's chorea: A demographic genetic and clinical study.* MD Thesis, University of London.

Stine, O., Pleasant, N., Franz, M., Abbott, M., Folstein, S. & Ross, C. (1993). Correlation between the onset age of Huntington's disease and length of the trinucleotide repeat in IT-15. *Human Molecular Genetics*, **2:10**, 1547–1549.

Trottier, Y., Biancalana, V., & Mandel, J.-L. (1994). Instability of CAG repeats in Huntingtons's disease: relation to parental transmission and age of onset. *Journal of Medicinal Genetics*, **31**, 377–382.

Venters, G. (1972). *Genetic and social aspects of Huntington's chorea.* PhD Thesis, University of Edinburgh.

Walker, D., Harper, P., Wells, C., Tyler, A., Davies, K. & Newcombe, R. (1981). Huntington chorea in South Wales. *Clinical Genetics*, **19**, 213–221.

Wallace, D.C. (1972). Huntington's chorea in Queensland. A not uncommon disease. *The Medical Journal of Australia*, **1**, 299–307.

Wilkie, A.D. (2000). *Report by the Independent Actuary on the Application by the Association of British Insurers to the Genetics and Insurance Committee for Approval to use Genetic Test Results for Insurance Risk Assessment — Use of Huntington's Disease Test in Life Insurance.* Report available by request to the Genetics and Insurance Committee.

Wendt, G.G. & Drohm, C. (1972). Die Huntingtonsche chorea. Eine populations gentische studie. *Advances in Human Genetics*, **IV**, 1–121.

Wendt, G.G., Landzettel, I. & Unterreiner, I. (1959). Erkrankungsalter bei der HuntingtonschenChorea. *Acta Genetica (Basel)*, **9**, 18–32.

Zühlke, C., Riess, O., Schroder, K., Siedlaczck, I., Epplen, J., Engel, W. & Thies, U. (1993). Expansion of the (CAG)n repeat causing Huntington's disease in 353 patients of German origin. *Human Molecular Genetics*, **2:9**, 1467–1469.