A NELSON-AALEN ESTIMATE OF THE INCIDENCE RATES OF EARLY-ONSET ALZHEIMER'S DISEASE ASSOCIATED WITH THE PRESENILIN-1 GENE

By Eng Hock Gui and Angus Macdonald

ABSTRACT

We analyse, in a probabilistic setting, Newcombe's (1981) life table method of estimating rates of onset of high-penetrance single-gene disorders, and extend this to a counting process model for individual life histories, including movement between risk groups arising from genetic testing and onset in relatives. We derive Nelson-Aalen-type estimates of the integrated intensity of onset, including a useful diagnostic check on the treatment of censored observations. We summarise the literature on mutations in the Presenilin-1 (PSEN-1) gene, associated with early-onset Alzheimer's disease (EOAD), and use published pedigrees to estimate (approximately) the rate of onset of EOAD.

KEYWORDS

Early-Onset Alzheimer's Disease; Nelson-Aalen Estimate; Presenilin-1 Gene; Pedigrees

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

1.1 *Estimating Rates of Onset of Genetic Disorders*

Among the human diseases caused by single-gene disorders are some whose sporadic occurrence (that is, in the absence of a gene mutation) is rare or unknown. In these cases, age-related rates of onset can be estimated using the life table method of Newcombe (1981), following Elandt-Johnson (1973). Before the introduction of DNA-based genetic testing, members of affected families could only be confirmed as mutation carriers once onset had occurred. The longer someone survived disease-free, therefore, the more likely it was that they did not carry the mutation, and observation of non-carriers would inevitably be censored. Some early estimates of rates of onset simply excluded these censored cases, but Newcombe (1981) pointed out that this was biased; his method properly allowed for them. It was obtained in a deterministic setting, in which the properties of the resulting estimates were not directly available.

If a (reliable) genetic test is available, asymptomatic members of affected families can be: untested and at risk of being a carrier; tested and known to be a carrier; or tested and known not to be a carrier. One of the aims of this paper is to extend Newcombe's method to allow for these new risk groups.

In Section 2, we describe pedigrees, or family histories, which are the data from which rates of onset can be estimated. An important question is how to make use of information

gained about risk status; this may change throughout a person's lifetime as their own and their relatives' life histories reveal information. A key feature of Newcombe's method is that all the information available at the time of the study is used to assign people to the same risk groups throughout their lives; in effect, making the best use of the latest available information.

In Section 3 we use a simple probabilistic model to study how information acquired at different times might be used in moment estimates of rates of onset, *via* conditional expectations or probabilities of onset. We show that, perhaps surprisingly, information acquired at any age x should not be used in estimating rates of onset at earlier ages. Intuitively, it might seem that its use should lead to better estimates, but in a probabilistic setting this is seen to be untrue.

Therefore, when there are a number of distinct risk groups, individuals may have to be assigned to different groups at different times, depending on the information that has by then emerged. In Section 4 we specify a model that allows for this, in a discrete-state continuous-time setting, and obtain Nelson-Aalen type estimates for certain functions of the transition intensity of interest, along with approximate confidence intervals. Further, in the simplest case we find that the function estimated is bounded above, unlike the usual integrated hazard, providing a diagnostic check for the inclusion of censored observations.

1.2 *Early-Onset Alzheimer's Disease*

For genetic terminology, see a text such as Pasternak (1999) or Strachan & Read (1999), or an introduction for non-specialists such as Fischer & Berberich (1999).

Alzheimer's disease (AD) is the most common cause of dementia, and a major health problem in many countries. It is characterized by a gradual and progressive decline in cognitive function. Early-onset Alzheimer's disease (EOAD) is AD occurring before about age 65. Whereas AD after that age is relatively common, EOAD is rare. Sometimes it runs in families, with a pattern of inheritance suggesting that one or more autosomal dominant genes are responsible. Three genes, amyloid precursor protein (APP), presenilin-1 (PSEN-1) and presenilin-2 (PSEN-2) have been confirmed as causing EOAD.

A fourth gene, ApoE, is associated with increased susceptibility to AD at older ages. Its implications for long-term care insurance or costs have been discussed by Macdonald & Pritchard (1999, 2000) and Warren *et al.* (1999). Other genes involved in AD, possibly in combination with ApoE, remain to be found.

APP and PSEN-1 gene mutations account for 10–15% and 20–70% of EOAD, respectively (Campion *et al.*, 1999). They are highly penetrant; the absence of AD by age 60 among confirmed carriers is rare. PSEN-2 gene mutations are very rare. The ages at onset of EOAD also vary (PSEN-1 25–60, APP 40–65 and PSEN-2 45–84 (Campion *et al.*, 1999)).

Pathogenic mutations of these genes are usually missense mutations, in which an error at a single DNA base causes an incorrect amino acid to be substituted in the protein produced by the gene. These genes are involved in producing the β amyloid peptide, that has 40 or more amino acids, and is found in the amyloid plaques in the brains of both sporadic and familial AD patients. The dominant hypothesis for AD (the 'amyloid cascade hypothesis') suggests that overproduction of, or failure to clear, the long forms of the peptide is crucial (Funato *et al.*, 1999; Hardy, 1997). Mutated genes produce Aβ_{42} or

$A\beta_{43}$ rather than the less pathogenic $A\beta_{40}$ (Younkin *et al.*, 1996; Czech, Tremp & Pradier, 1999). See St. George-Hyslop (2000) for a review of the molecular genetics of AD.

Dartigues & Letenneur (2000) describe the genetic epidemiology of AD, which is almost entirely related to the ApoE gene. No estimates of incidence rates of EOAD have been published, and very few estimates of prevalence; little is known about mutation frequencies. The estimates of incidence rates in respect of PSEN-1 mutations in this paper are novel.

1.3 *The Presenilin-1 Gene*

Also known as S182, the PSEN-1 gene was localised on chromosome 14 (location 14q24) in 1992 but could only be isolated in 1995 (Sherrington et al, 1995). It is highly conserved in evolution, so mutations are rare. The coding region comprises 10 exons, numbered 3–12, and encodes a transmembrane protein that is produced at low level in many different cell types, and is almost homogeneously expressed in the brain and in peripheral tissues.

To date, nearly 100 different mutations causing EOAD have been found. PSEN-1 mutations are usually associated with very aggressive EOAD, with duration of dementia of about 5 years (Russo *et al.*, 2000) Appendix A summarises the families in which PSEN-1 mutations have been found. The lowest ages at onset are around 30, while there are rare survivors with a mutation at sufficiently high ages to suggest incomplete penetrance. The few isolated cases could be sporadic rather than inherited mutations. Overall, however, mutations appear to be highly penetrant before the age of about 60.

A striking feature is that ages at onset tend to be similar within families, but different in different families. As far as insurance risk is concerned, this suggests that a genetic test result that shows a mutation to be present should be significantly more informative if the ages of affected relatives are known. However, for studying aggregated or average costs (for example, the economic costs of long-term care or the costs of adverse selection in insurance) we would need to model explicitly how familial clustering is distributed in the whole population. For example, if we chose the mean age at onset of AD within a family as a covariate, we would have to estimate its distribution in the population. Estimating the rate of onset of EOAD from the whole sample, ignoring familial clustering, achieves the same aim more simply.

1.4 *Estimating Rates of Onset of Early-Onset Alzheimer's Disease*

The second aim of this paper is to obtain age-related estimates of the rates of onset of EOAD associated with the PSEN-1 gene. For actuarial applications, we think in terms of a continuous-time, finite-state stochastic process model (see Macdonald, 1997, 1999; Subramanian *et al.*, 2000) in which some of the transitions represent onset of AD; the rates of onset are then the transition intensities in the model.

Many of the articles on PSEN-1 mutations (summarised in Appendix A) include pedigrees, from which rates of onset can, in theory, be estimated, but few give all the times of entry to and exit from all relevant risk groups, especially for censored cases, so we have had to make some approximations, described in Section 5.2. The details of the estimation are in Sections 5.3 to 5.5, and conclusions are in Section 6.

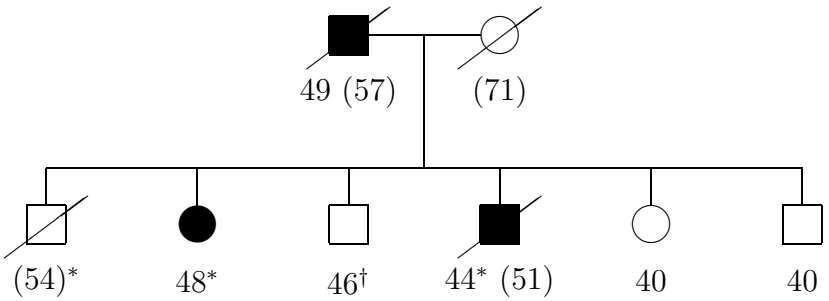

Figure 1: A hypothetical example of a pedigree. Squares are males, circles are females, and a slash denotes death. Affected individuals are shown as filled squares/circles. The age at onset or oldest observed age free of AD is shown, and age at death is given in brackets. An asterisk means that a person has been tested and does carry a mutation, a dagger that he/she has been tested and does not carry a mutation. By convention siblings are listed left-to-right in birth order.

2. Pedigree Analysis

2.1 *Pedigrees and Mutation Probabilities*

Figure 1 gives a hypothetical example of a pedigree. The father in this example did not have a genetic test, but EOAD is so rare that we can reasonably assume that he did carry the mutation. Of the second generation, the first, second and fourth lives have been tested and have a mutation; the third life has been tested and has no mutation; and the fifth and sixth lives have not been tested and are at risk of having the mutation. The probabilities that they have a mutation depend on the available information. By Mendel's laws, these probabilities would appear to be 1/2 each, but this may have to be modified in complicated ways. For example, denote the father F and the fifth child C.

(a) Survival free of EOAD reduces the probability that someone at risk has a mutation, ultimately to zero if the mutation is fully penetrant. Let $p(x)$ be the probability that a mutation carrier is free of EOAD at age x. Then the probability that an at-risk individual who is free of EOAD at age x is a carrier (ignoring all other decrements) is $p(x)/(1+p(x))$, which is the Mendelian 1/2 at age 0.

(b) F's carrier status was not known when C was born; only when his condition appeared later on. If one of F's parents had had EOAD then, when C was born, the Mendelian probability of her having a mutation would be 1/4, not 1/2, but again this would be modified by survival. For example, if C was born when F was age 25, C's mutation probability at age x, while F was healthy (again ignoring all other decrements), would be:

$$\frac{0.25p(x)p(x+25)}{0.5+0.25p(x+25)+0.25p(x)p(x+25)}. \tag{1}$$

(c) Even this may fail to use all the available information. While F is healthy, the survival of C's siblings free of symptoms (and before having had a genetic test) also decreases

the probability that C carries a mutation. See Newcombe (1981) for an example of these calculations.

Complications (b) and (c) above arise only when the carrier status of F and all of C's siblings is uncertain. As soon as any one of them develops EOAD, and can be assumed to have the mutation, C's mutation probability reverts to $p(x)/(1+p(x))$. Alternatively, if C were to have a (100% accurate) genetic test, her mutation probability would then change to 0 or 1. Similarly, as long as F remains free of EOAD, his mutation probability depends upon the continuing freedom from EOAD of all his children.

We ignored other decrements above (death from other causes or loss to follow-up) which would, in practice, introduce censoring mechanisms.

2.2 *Definition of Risk Groups*

At any time, a family member is in one of four risk groups:

(a) *Group 1: Mutation known to be absent.* As a result of a genetic test, taken by the person or an ancestor, the mutation probability is 0 (assuming the test to be accurate).

(b) *Group 2: Mutation known to be present.* Either a genetic test has revealed a mutation, or EOAD has occurred. EOAD is rare enough that, within affected families, its presence may be assumed to indicate the presence of a gene mutation, though strictly a genetic test would be necessary to confirm it. This assumption could not be made in the case of a condition that commonly occurred for other reasons (for example, breast cancer).

(c) *Group 3: Parent or any sibling known to have the mutation.* The mutation probability is $p(x)/(1+p(x))$. This depends only on the age of the individual concerned, it does not depend explicitly on the details of the family.

(d) *Group 4: Parent and siblings at risk but free of EOAD.* The mutation probability is complicated and depends explicitly on the details of the family; Equation (1) is the simplest case.

We wish to estimate age-related rates of onset of EOAD. It is clear that the risk groups defined above will enter the analysis somehow. For example, consider simple occurrence/exposure rates. The numerators present no problems, but the exposures will have to allow for the different risk groups. An intuitive approach is to weight each person's contribution to the exposure at age x by the probability that they have the mutation. Strictly, we should say the *conditional* probability that they have the mutation, since it depends on the data. A key question is this: for a person at risk at age x, what information should we use to condition the probability that they have the mutation?

An example will show why the question arises. Consider an only child who is age 10 when her mother suffers EOAD, and who suffers it herself at age 40, dying at age 50. How does she contribute to the exposure at age 20? We could use this life history in two ways:

(a) She was in risk group 4 for ten years, risk group 3 for thirty years and risk group 2 for ten years. At age 20, *based only on what was known at that time* she had the mutation with probability $p(20)/(1+p(20))$, which is therefore her contribution to the exposure.

(b) By the time the analysis is carried out, after her death, it is known that she had had EOAD so she was, in fact, in risk group 2 throughout her life. She therefore contributes one full year to the exposure at age 20.

The latter is the basis of the 'life table method' of Newcombe (1981). Intuitively, it seems sensible to condition on all the information available at the time of the study, and to use this 'best knowledge' to assign each person to one risk group throughout their life. However, this leads to problems in interpreting the resulting occurrence/exposure rates.

2.3 *The Life Table Method*

Based on Elandt-Johnson (1973), Newcombe (1981) introduced a modified life table analysis for Huntington's disease (HD) to allow for risk pools like Groups 3 and 4 above (this work predated the discovery of the HD gene). Previous analyses had estimated occurrence/exposure rates based only on affected persons (risk group 2 above) which were clearly biased, especially at older ages, because the omission of censored observations understated the true numbers at risk. Newcombe's estimate added terms to the exposure to allow for censored observations of at-risk individuals. Harper & Newcombe (1992) tabulated probabilities of being a carrier, given age and family history, that are used in counselling for people at risk of HD.

Four groups were distinguished; obligate carriers with mutation probability 1, and at-risk individuals with Mendelian mutation probability 0.5, 0.25 or 0.125, depending on the last generation affected. Suppose that A_x persons suffered onset of HD at age x, and observation of N_x lifetimes was censored at age x. For each of these, the mutation probability was calculated allowing for all the information available at the time of the study, including what was known about relatives; call this R_x^i in respect of the i^{th} life censored at age x. The exposure to risk, during each full year of observation, in respect of lives censored at age x was taken to be $\sum_{i=1}^{i=N_x} R_x^i$, and the effective total (initial-type) exposure at age x was:

$$E_x = \sum_{y \geq x} A_y + \sum_{y > x} \sum_{i=1}^{i=N_y} R_y^i + \frac{1}{2} \sum_{i=1}^{i=N_x} R_x^i. \tag{2}$$

Because of the appearance of rates of onset in the R_x^i, it was impossible to estimate one-year probabilities as A_x/E_x in the usual way; the system of 71 non-linear equations was solved by iteration.

This method seems intuitively reasonable, but it was not derived from any underlying probabilistic model, and its properties are hard to find for that reason as well as because of its inherent complexity. Confidence limits were not available, for example. A simulation experiment showed that the mean age at onset was biased slightly upwards; the bias was statistically but not practically significant.

3. An Analysis of the Life Table Method

Occurrence/exposure rates arise very naturally as moment estimates in the context of probabilistic models for censored data. Put simply, we proceed from the model specification to an equation of the form:

$$\text{E[No. of Events} \mid \text{Information]} = \text{P[Event]} \times E \tag{3}$$

where E is some quantity that is known from the given information. As the notation suggests, E can often be interpreted as an intuitive 'exposure to risk', hence the name 'occurrence/exposure rates' for the resulting moment estimates. In most simple cases, moment and likelihood estimates coincide (we concentrate on moment estimates, because that is the simplest interpretation of the Nelson-Aalen estimate we use later).

In Elandt-Johnson (1973) and Newcombe (1981) no probabilistic model was specified explicitly, and expected values were implicitly replaced by proportions. Here we use a simple probabilistic model, that includes the key features of Newcombe's estimation procedure, to pose some questions about the conditional expectations that may underlie Equations (2) and (3).

Suppose we begin with a sample of N independent lives who, at birth, each have a mutation with probability 1/2. A person with the mutation, healthy at age x, has probability q_x of suffering onset of the disorder by age $x+1$. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be the number of cases of onset during the first year of life, second year of life and so on. We observe all cases of onset until some fixed age T, when observation ceases. Onset of the disorder is the only decrement, so the only cause of censoring is by remaining unaffected at age T. The problem is to estimate the q_x at integer ages x.

3.1 *Conditioning on Currently Known Information*

First, we can write down the conditional expectations, assuming only that we know what happened up to the time (or age) for which a rate is being estimated; we denote information known at age x by $\mathcal{F}_x$:

$$\begin{aligned}
\mathrm{E}[\mathbf{X}_1|\mathcal{F}_0] &= \frac{1}{2} N q_0 && (4)\\
\mathrm{E}[\mathbf{X}_2|\mathcal{F}_1] &= (N - \mathbf{X}_1)\frac{(1-q_0)}{1+(1-q_0)} q_1 && (5)\\
\mathrm{E}[\mathbf{X}_3|\mathcal{F}_2] &= (N - \mathbf{X}_1 - \mathbf{X}_2)\frac{(1-q_0)(1-q_1)}{1+(1-q_0)(1-q_1)} q_2 \quad \text{etc.} && (6)\\
& && (7)
\end{aligned}$$

This is exactly what would be done in a conventional survival analysis. Censored cases and cases of onset alike are included in the exposure, and there is no bias arising from the omission of censored cases.

3.2 *Conditioning Retrospectively on Observed Cases*

Newcombe (1981) takes advantage of the fact that the analysis is retrospective, and is carried out at time T. By that time, $\mathbf{X}^* = \mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_T$ individuals have been identified as mutation carriers, and $(N - \mathbf{X}^*)$ may or may not be. This information is used to split the sample population in two throughout the analysis:
(a) the $\mathbf{X}^*$ known to be mutation carriers; and

(b) the $N - \mathbf{X}^*$ not known to be mutation carriers, each of whom is a mutation carrier with probability $P/(1+P)$, where:

$$P = (1-q_0)(1-q_1)\dots(1-q_{T-1}). \tag{8}$$

Note that knowing $\mathbf{X}^*$ is not the same as knowing $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T$ separately.

Conditional expected values of $\mathbf{X}_1, \mathbf{X}_2, \dots$ are then found using this additional information; here we denote the information at age x by $\mathcal{F}_x^*$. Note that $\mathrm{E}[\mathbf{X}^*|\mathcal{F}_0] = N(1-P)/2$, while $\mathrm{E}[\mathbf{X}^*|\mathcal{F}_0^*] = \mathbf{X}^*$.

$$\begin{aligned}
\mathrm{E}[\mathbf{X}_1|\mathcal{F}_0^*] &= \mathbf{X}^* \frac{q_0}{1-(1-q_0)(1-q_1)\dots(1-q_{T-1})} && (9)\\
\mathrm{E}[\mathbf{X}_2|\mathcal{F}_1^*] &= (\mathbf{X}^* - \mathbf{X}_1) \frac{q_1}{1-(1-q_1)(1-q_2)\dots(1-q_{T-1})} && (10)\\
\mathrm{E}[\mathbf{X}_3|\mathcal{F}_2^*] &= (\mathbf{X}^* - \mathbf{X}_1 - \mathbf{X}_2) \frac{q_2}{1-(1-q_2)(1-q_3)\dots(1-q_{T-1})} \quad \text{etc.} && (11)\\
& && (12)
\end{aligned}$$

We derive Equation (10) as an example. Let $\mathbf{X}_j^i$ be the indicator of onset of the disorder during the j^{th} year of life of the i^{th} person. That is, $\mathbf{X}_j^i = 1$ if the i^{th} person suffers onset between ages $j-1$ and j, and $\mathbf{X}_j^i = 0$ otherwise. Then $\mathbf{X}_j = \sum_{i=1}^{i=N} \mathbf{X}_j^i$. Further, define $\mathcal{C}$ to be the set of indices corresponding to the $\mathbf{X}^*$ identified mutation carriers, and let $\mathcal{U}$ be the set of indices corresponding to censored observations. That is, $i \in \mathcal{C}$ if the i^{th} life suffered onset, and $i \in \mathcal{U}$ otherwise. Then:

$$\begin{aligned}
\mathrm{E}[\mathbf{X}_2|\mathcal{F}_1^*] &= \sum_{i=1}^{i=N} \mathrm{E}[\mathbf{X}_2^i|\mathcal{F}_1^*] && (13)\\
&= \sum_{i=1}^{i=N} \mathrm{P}[\mathbf{X}_2^i = 1|\mathcal{F}_1^*] && (14)\\
&= \sum_{\substack{i\in\mathcal{C}\\ \mathbf{X}_1^i=0}} \mathrm{P}[\mathbf{X}_2^i = 1 | i \in \mathcal{C}, \mathbf{X}_1^i = 0] + \sum_{\substack{i\in\mathcal{U}\\ \mathbf{X}_1^i=0}} \mathrm{P}[\mathbf{X}_2^i = 1 | i \in \mathcal{U}, \mathbf{X}_1^i = 0] && (15)\\
&= (\mathbf{X}^* - \mathbf{X}_1) \frac{q_1}{1-(1-q_1)(1-q_2)\dots(1-q_{T-1})}. && (16)
\end{aligned}$$

In going from Equation (14) to Equation (15) we split the summation between sets of indices that are known given $\mathcal{F}_1^*$, and then the second sum clearly disappears because $\mathrm{P}[\mathbf{X}_2^i = 1 | i \in \mathcal{U}, \mathbf{X}_1^i = 0] = \mathrm{P}[\mathbf{X}_2^i = 1 | i \in \mathcal{U}] = 0$, by definition of $\mathcal{U}$.

This approach apparently yields a system of T equations in T unknowns, which we might try to solve. Unfortunately $\mathrm{E}[\mathbf{X}_T|\mathcal{F}_{T-1}^*] = \mathbf{X}_T$, so we must have $q_{T-1} = 1$ and the solution collapses. Essentially, this is because we have conditioned on the very statistic we need for estimation. Any attempt to write down a conditional likelihood, conditioning on $\mathcal{C}$ and $\mathcal{U}$, will suffer similar problems.

Note that these estimates are the same as those obtained by ignoring censored observations, that Newcombe (1981) observed to be biased.

3.3 *Newcombe's Estimates*

The exposures in Newcombe's estimates are different; they are:

$$E_0 = \mathbf{X}^* + (N - \mathbf{X}^*)\frac{P}{1+P} \tag{17}$$

$$E_1 = (\mathbf{X}^* - \mathbf{X}_1) + (N - \mathbf{X}^*)\frac{P}{1+P} \tag{18}$$

$$E_2 = (\mathbf{X}^* - \mathbf{X}_1 - \mathbf{X}_2) + (N - \mathbf{X}^*)\frac{P}{1+P} \quad \text{etc.} \tag{19}$$

$$\tag{20}$$

These correspond to Equations (2), except that here, censoring occurs only at one fixed age. As can be seen from Section 3.2, they are not the conditional expectations, given the information used by Newcombe (1981), and the resulting estimates are not moment estimates (or conditional likelihood estimates). Nevertheless, they are consistent with moment estimates in the sense that $\mathrm{E}[E_0|\mathcal{F}_0]q_0 = \mathrm{E}[\mathbf{X}_1|\mathcal{F}_0]$, $\mathrm{E}[E_1|\mathcal{F}_1]q_1 = \mathrm{E}[\mathbf{X}_2|\mathcal{F}_1]$ and so on (as can easily be checked) so for reasonably large samples, they should be quite similar. In a probabilistic framework, however, conditioning on information that includes the very events being studied fails. The absence of a probabilistic framework also means that the properties of the Newcombe's estimator cannot be studied directly.

3.4 *Information From Events Other Than Onset*

Information about a person's genotype can be revealed by events other than they themselves suffering onset of the disorder. Onset in relatives, or genetic tests taken by the person or their relatives, can alter their risk status. How should people be assigned to risk groups in light of these events? For example:

(a) Consider identical twin brothers, at risk 1/2 of being carriers. One suffers onset at age 40. As we saw in Section 3.2, we cannot then assign him to the 'known carrier' risk group at earlier ages, but can we so assign his brother?

(b) Consider someone age 30 who takes a genetic test that shows they are a carrier. This is not the same event as that whose rate of occurrence we wish to estimate, so can we assign that person to the 'known carrier' risk group at earlier ages? And can we use this information when assigning their relatives (for example, children) to risk groups?

It should be clear that the answer to (a) above is no. The event whose rate of occurrence is being estimated cannot appear in the information set used in conditional expectations or probabilities. This is perhaps most obvious if we consider the likelihood, which would be the conditional *joint* probability of the observed events befalling both brothers; we cannot condition on different information in respect of each.

To explore (b) above we adapt our simple model, and suppose that at time T, $\mathbf{X}^\dagger$ asymptomatic people have a genetic test. This creates four subgroups: $\mathcal{C}$ as before, and $\mathcal{P}, \mathcal{N}$ and $\mathcal{U}^*$, respectively those who are tested with positive result (mutation present),

those who are tested with negative result, and those who are not tested. Assume genetic test results are included in the conditioning at all ages, but later cases of onset are not, and denote the information at age x by $\mathcal{F}_x^\dagger$. Then:

$$\begin{aligned}
\mathrm{E}[\mathbf{X}_1|\mathcal{F}_0^\dagger] &= \frac{1}{2}(N - \mathbf{X}^\dagger)q_0 && (21)\\
\mathrm{E}[\mathbf{X}_2|\mathcal{F}_1^\dagger] &= (N - \mathbf{X}_1 - \mathbf{X}^\dagger)\frac{(1-q_0)}{1+(1-q_0)}q_1 && (22)\\
\mathrm{E}[\mathbf{X}_3|\mathcal{F}_2^\dagger] &= (N - \mathbf{X}_1 - \mathbf{X}_2 - \mathbf{X}^\dagger)\frac{(1-q_0)(1-q_1)}{1+(1-q_0)(1-q_1)}q_2 \quad \text{etc.} && (23)\\
& && (24)
\end{aligned}$$

The reasoning is the same as before: if someone is in $\mathcal{P}$ or $\mathcal{N}$, they are known to be asymptomatic at age T, so the conditional probability of onset at any earlier age is zero.

We consider one more possibility: if the problem is that the age at which a genetic test was taken is known, can we not simply 'forget' that piece of information, and then assign tested individuals to a 'known status' risk group throughout their lives? Denote this information at age x by $\mathcal{F}_x^\ddagger$ and consider (for example):

$$\mathrm{E}[\mathbf{X}_2|\mathcal{F}_1^\ddagger] = \sum_{i=1}^{i=N} \mathrm{P}[\mathbf{X}_2^i = 1|\mathcal{F}_1^\ddagger]. \tag{25}$$

Suppose $i \in \mathcal{P}$. This means that the i^{th} life is known to have had a positive test *before onset*, but when is unknown. Suppose the unknown time is $\mathbf{T}^i$. Then:

$$\mathrm{P}[\mathbf{X}_2^i = 1|\mathcal{F}_1^\ddagger] = \mathrm{P}[\mathbf{T}^i \leq 1]q_1 + \mathrm{P}[\mathbf{T}^i > 1] \times 0. \tag{26}$$

and:

(a) we have effectively removed a proportion of the at-risk population, leading to the same sort of biased estimate as if only observed cases of onset were included; and
(b) we have had to introduce age-related probabilities of taking a genetic test.

In a probabilistic setting, every attempt to use knowledge obtained at future ages to improve estimates of rates of onset at past ages is foiled because, conditioning on that information, we know that onset did not occur. And, although we have considered conditional expectations and moment estimates, the same will be true of conditional probabilities and likelihood estimates.

In this section we have omitted any complications, but it is clear that we should not assign individuals retrospectively to one risk group throughout their lives, but allow them to be in different risk groups at different ages. This leads naturally to continuous-time discrete-state stochastic process models. In Section 4.1 we will specify such a model in which a person age x is assigned to a risk group using only the information available at age x, and we will derive a modified version of the Nelson-Aalen estimate.

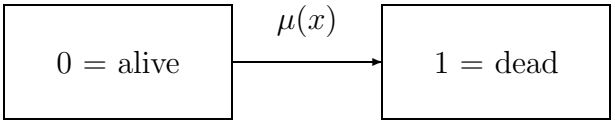

Figure 2: A two state model of mortality

3.5 *A Significant Corollary*

This conclusion has a significant corollary, concerning the information that ought to be included in pedigrees. A person moves from one risk group to another either when they have a genetic test, or when a relative has a genetic test or suffers onset of EOAD. Therefore:

(a) the age at which an asymptomatic person has a genetic test; and
(b) the affected parent's age at the birth of each child

must be recorded as part of the pedigree; it is not enough just to know a test result, or the unconnected ages of family members when events befall them.

4. A Modified Nelson-Aalen Estimate

From now on, we will often refer to the rate of onset as a transition intensity (or just intensity), denoted $\mu(x)$ as a function of age x, adding subscripts in models involving more than one intensity. It is a natural target for estimation because it is a key quantity in many actuarial applications. Equivalently, we can estimate any simple function of $\mu(x)$. One such function is the integrated intensity $\int_0^x \mu(t)dt$, for which there is a natural estimate, the Nelson-Aalen estimate. This has nice statistical properties, but for our purposes its advantage is that it can be generalised to allow for uncertainty about genotype.

In Section 4.1 we introduce the Nelson-Aalen estimate, then we modify it in stages:

(a) in Sections 4.2 and 4.3 we allow for an unaffected person with Mendelian probability p at birth of having a mutation;
(b) in Section 4.4 we allow for genetic testing; and
(c) in Section 4.5 we allow for (or rather, are defeated by) the information contributed by unaffected relatives.

4.1 *The Nelson-Aalen Estimate*

Here we introduce, non-rigorously, the Nelson-Aalen estimate. A fuller, but still heuristic, introduction to counting processes can be found in Macdonald (1996); see Fleming & Harrington (1991) or Andersen *et al.* (1993) for a proper treatment.

Suppose we observe M lives, each of whose survival can be represented by the continuous-time model in Figure 2. The state space is $\mathcal{S} = \{0, 1\}$, and at age x the state occupied by the i^{th} person is denoted $\mathbf{S}^i(x)$, $\mathbf{N}^i(x)$ is the number of transitions from alive to dead by age x (either 0 or 1), and $\mathbf{Y}^i(x)$ indicates presence in the alive state just before age x ($\mathbf{Y}^i(x) = 1$ if the life is alive just before age x, and is 0 otherwise). $F^i = \{\mathcal{F}_x^i\}_{x\geq 0}$ is

the natural filtration of $\mathbf{N}^i(x)$. The compensated process $\mathbf{M}^i(x) = \mathbf{N}^i(x) - \int_0^x \mathbf{Y}^i(t)\mu(t)dt$ is a F^i-martingale, and $\int_0^x \mathbf{Y}^i(t)\mu(t)dt$ is the F^i-compensator of $\mathbf{N}^i(x)$.

Sum these quantities over the M lives ($\mathbf{N}(x) = \sum_{i=1}^M \mathbf{N}^i(x)$ and so on, absence of the superscript i denoting such sums) and let $F = \{\mathcal{F}_x\}_{x\geq 0}$ be the natural filtration. $\mathbf{M}(x)$ is a F-martingale and, if $\mathbf{H}(x)$ is a predictable process adapted to F, the stochastic integral $\int_0^x \mathbf{H}(t)d\mathbf{M}(t) = \int_0^x \mathbf{H}(t)d\mathbf{N}(t) - \int_0^x \mathbf{H}(t)\mathbf{Y}(t)\mu(t)dt$ is also a F-martingale, zero at $x = 0$. Define $\mathbf{J}(x) = \mathbf{I}_{\{\mathbf{Y}(x)>0\}}$, with the convention that $\mathbf{Y}(x) = 0 \Rightarrow \mathbf{J}(x)/\mathbf{Y}(x) = 0$, and take $\mathbf{H}(x) = \mathbf{J}(x)/\mathbf{Y}(x)$. Then:

$$\mathrm{E}\left[\int_0^x \frac{\mathbf{J}(t)}{\mathbf{Y}(t)} d\mathbf{N}(t)\right] = \mathrm{E}\left[\int_0^x \mathbf{J}(t)\mu(t)dt\right] \approx \int_0^x \mu(t)dt. \tag{27}$$

$\hat{\Lambda}(x) = \int_0^x \mathbf{J}(t)\mathbf{Y}^{-1}(t)d\mathbf{N}(t)$ is the Nelson-Aalen estimate of $\int_0^x \mu(t)dt$. In words: at age x, $\mathbf{Y}(x)$ lives are at risk. If one dies, the estimate increases by $1/\mathbf{Y}(x)$, provided $\mathbf{Y}(x) > 0$. $\mathbf{J}(x)$ takes care of the possibility that $\mathbf{Y}(x) = 0$. In between observed death times, the estimate is level. Equation (27) shows that $\hat{\Lambda}(x)$ is 'almost' unbiased, the bias arising from the possibility that no lives remain under observation.

The variance of the Nelson-Aalen estimate can be estimated reasonably well by:

$$\mathrm{Var}[\hat{\Lambda}] \approx \int_0^x \frac{\mathbf{J}(t)(\mathbf{Y}(t) - \Delta\mathbf{N}(t))}{(\mathbf{Y}(t))^3} d\mathbf{N}(t) \tag{28}$$

(Andersen *et al.*, 1993), where $\Delta\mathbf{N}(t)$ is the jump in $\mathbf{N}(t)$ at time t, leading to pointwise confidence intervals. However, in Section 4.2 we estimate a more complicated function of the intensity, and we will use the so-called 'Weird Bootstrap' (Andersen *et al.*, 1993).

(a) At each jump time of $\mathbf{N}(x)$, fix the numbers at risk at their observed values $\mathbf{Y}(x)$.
(b) Simulate the number of deaths at the jump times, as a Binomial$(\mathbf{Y}(x), d\mathbf{N}(x)/\mathbf{Y}(x))$ random variable.
(c) Calculate a simulated Nelson-Aalen estimate, and solve for the intensity if desired.
(d) Over many simulations, the distribution of $\int_0^x \hat{\mu}(t)dt$ or $\hat{\mu}(x)$ is built up, from which pointwise confidence intervals can be found directly.

The indicators $\mathbf{Y}^i(x)$ and $\mathbf{Y}(x)$ can allow for many censoring schemes. In this case the natural filtration is $\mathcal{F}_x^i = \sigma(\mathbf{N}^i(t), \mathbf{Y}^i(t) : t \leq x)$, but for simplicity we ignore such censoring in what follows.

4.2 *A Modified Nelson-Aalen Estimate*

Figure 3 shows a model representing the onset of AD, with two starting states: State 0 (has mutation) and State 1 (no mutation). The state space is now $\mathcal{S} = \{0, 1, 2\}$. Therefore:

(a) $\mu_{02}(x)$ is the incidence rate that we wish to estimate; and
(b) $\mu_{12}(x)$ is the incidence rate of sporadic EOAD, assumed known (often zero).

We suppose that the person has not had a genetic test, so we do not know in which state they start (at birth), but we know the Mendelian probability p that, at birth, they

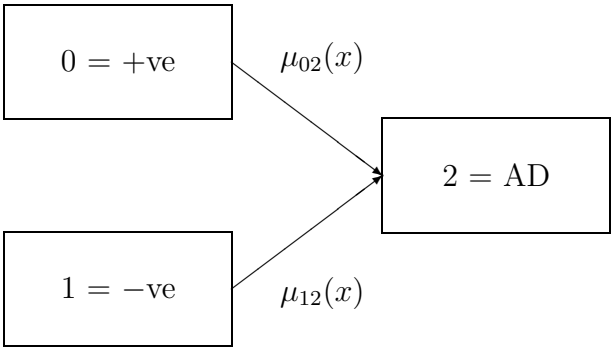

Figure 3: A model of the incidence of Alzheimer's disease where an individual may have an EOAD mutation (State 0, +ve) or may not have an EOAD mutation (State 1, −ve).

are in State 0, and $1-p$ that they are in State 1, for example because an ancestor is a known mutation carrier (Groups 3 and 4 in Section 2.2).

Here, we modify the Nelson-Aalen estimate, and show that it estimates a function of $\mu_{02}(x)$, more complicated than $\int_0^x \mu_{02}(t)dt$, but capable of numerical solution.

In obvious notation, the model for the i^{th} life history is the multivariate counting process $(\mathbf{N}^i_{02}(x), \mathbf{N}^i_{12}(x))$ and indicators $(\mathbf{Y}^i_0(x), \mathbf{Y}^i_1(x))$. However, we cannot observe these separately, but only $\mathbf{N}^i(x) = \mathbf{N}^i_{02}(x) + \mathbf{N}^i_{12}(x)$ and $\mathbf{Y}^i(x) = \mathbf{Y}^i_0(x) + \mathbf{Y}^i_1(x)$, so our information is:

(a) p, the Mendelian probability of carrying a mutation; and
(b) the filtration $G^i = \{\mathcal{G}^i_x\}_{x\geq 0}$, where $\mathcal{G}^i_x = \sigma(\mathbf{N}^i(t) : t \leq x)$; we observe whether or not AD has appeared by age x.

It is easily checked that:

$$\mathbf{A}^i_{\mathcal{G}}(x) = \int_0^x \mathbf{Y}^i(t) \frac{p\exp(-\int_0^t \mu_{02}(s)ds)\mu_{02}(t) + (1-p)\exp(-\int_0^t \mu_{12}(s)ds)\mu_{12}(t)}{p\exp(-\int_0^t \mu_{02}(s)ds) + (1-p)\exp(-\int_0^t \mu_{12}(s)ds)} dt \quad (29)$$

is the G^i-compensator of $\mathbf{N}^i(x)$. Since $\mu_{02}(x)$ represents a highly penetrant disorder, and $\mu_{12}(x)$ a rare sporadic event, $\mu_{02}(x) \gg \mu_{12}(x)$, and we approximate $\mu_{12}(x) \approx 0$. Then:

$$\mathbf{A}^i_{\mathcal{G}}(x) \approx \int_0^x \mathbf{Y}^i(t) \frac{p\exp(-\int_0^t \mu_{02}(s)ds)\mu_{02}(t)}{p\exp(-\int_0^t \mu_{02}(s)ds) + (1-p)} dt = \int_0^x \mathbf{Y}^i(t)\lambda(t,\mu)\mu_{02}(t)dt \quad (30)$$

which we take as the definition of $\lambda(t,\mu)$. (We define $\lambda(t,\mu)$ for notational convenience only; it is a function of $\mu_{02}(s)$ for all $s \leq t$.) Now sum over all lives ($\mathbf{N}(x) = \sum_{i=1}^M \mathbf{N}^i(x)$ and so on) and define $\mathbf{J}(x) = \mathbf{I}_{\{\mathbf{Y}(x)>0\}}$ as before, then:

$$\hat{\Lambda}(x) = \int_0^x \frac{\mathbf{J}(t)}{\mathbf{Y}(t)} d\mathbf{N}(t) \quad (31)$$

is an 'almost' unbiased estimate of $\int_0^x \lambda(t,\mu)\mu_{02}(t)dt$. The obvious procedure is then: estimate $\hat{\Lambda}(x)$, let $\tilde{\Lambda}(x)$ be a smoothed version of it, and estimate $\hat{\mu}_{02}(x)$ by solving:

$$\frac{d\tilde{\Lambda}(x)}{dx} = \lambda(x,\mu)\mu_{02}(x). \tag{32}$$

numerically. Alternatively, note that:

$$\hat{\Gamma}(x) = \int_0^x \frac{\mathbf{J}(t)}{\mathbf{Y}(t)\lambda(t,\mu)} d\mathbf{N}(t) \tag{33}$$

estimates $\int_0^x \mu_{02}(t)dt$, which we could solve by iteration (rather as Newcombe (1981) did). We may take:

$$\int_0^x \frac{\mathbf{J}(t)(\mathbf{Y}(t)-\Delta\mathbf{N}(t))}{(\mathbf{Y}(t))^3} d\mathbf{N}(t) \qquad \text{and} \qquad \int_0^x \frac{\mathbf{J}(t)(\mathbf{Y}(t)-\Delta\mathbf{N}(t))}{\lambda(x,\mu)^2(\mathbf{Y}(t))^3} d\mathbf{N}(t) \tag{34}$$

to estimate $\mathrm{Var}[\hat{\Lambda}(x)]$ and $\mathrm{Var}[\hat{\Gamma}(x)]$, respectively (see Andersen *et al.* (1993) Section IV.1). The 'Weird Bootstrap' (Section 4.1) can be used exactly as before to find confidence intervals for $\mu_{02}(x)$; at each event time we have a risk set containing an unknown true number of lives with a mutation, but by using the same Binomial distribution we correctly re-sample from the observations.

4.3 *A Diagnostic Check for the Inclusion of Censored Observations*

Let $\Lambda(x)$ be the solution of Equation (32), based on the 'true' intensity $\mu_{02}(x)$. The 'true' version of Equation (32) can be put in the form:

$$\frac{d}{dx}f(x) + c(x)f(x) = \frac{p-1}{p}c(x) \tag{35}$$

where $c(x) = \Lambda'(x)$ and $f(x) = \exp(-\int_0^x \mu_{02}(t)dt)$. Solving this linear ODE, we see that any solution of Equation (32) with $f(0) = 1$ and $\Lambda(0) = 0$ satisfies:

$$\exp\left(-\int_0^x \mu_{02}(t)dt\right) = \frac{(1-p)^{-1} - e^{\Lambda(x)}}{p(1-p)^{-1}e^{\Lambda(x)}} \tag{36}$$

which means that $\Lambda(x)$ cannot exceed $-\log(1-p)$. In practice, $\hat{\Lambda}(x)$ can exceed $-\log(1-p)$, in which case $\hat{\mu}_{02}(x)$ explodes to infinity. If $p = 1/2$, $-\log(1-p)$ is only 0.693. Figure 4 shows the true $\Lambda(x)$, and six simulated estimates of $\hat{\Lambda}(x)$ between ages 0–60 (sample size 30, all observed from age 0, with random censoring taking place from age 20, heavier near the younger ages), for $p = 1/2$ and a hypothetical intensity $\mu_{02}(x) = (-x^4 + 85x^3 - 25x^2) \times 10^{-7}$ that reaches about 0.5 at age 60, so penetrance is close to 100%. Two of the simulated estimates exceed the bound. This is quite predictable at older ages, where the exposure is small; however the exposure may be understated for other reasons. If $\hat{\Lambda}(x)$ exceeds its bound at a rather early age, where exposures are still reasonable, we should suspect that some censored cases have been excluded from the data, perhaps because vital information such as age at censoring has not been recorded. This may provide a useful diagnostic check.

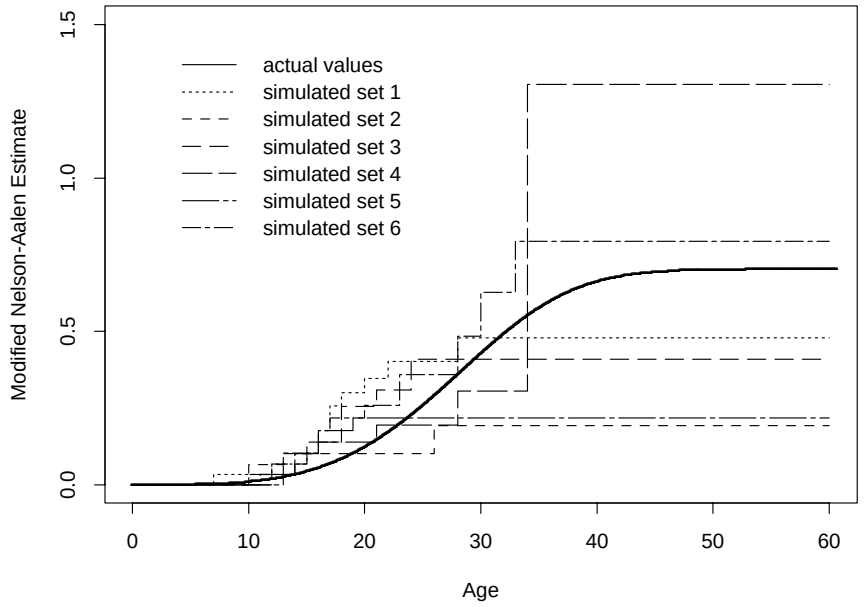

Figure 4: $\Lambda(x)$, and six simulated estimates of $\hat{\Lambda}(x)$ with sample size 30, for a hypothetical intensity $\mu_{02}(x) = (-x^4 + 85x^3 - 25x^2) \times 10^{-7}$ and $p = 1/2$, with random censoring.

Equation (36) extends to a heterogeneous model, in which different rates of onset may be associated with different mutations. Suppose there are two mutations, that an 'at-risk' person carries with probabilities p_1 and p_2, respectively ($p_1 + p_2 = p$). The associated rates of onset are $\mu^1_{02}(x)$ and $\mu^2_{02}(x)$, respectively. The form of $\Lambda(x)$ is found by modifying Equation (29) in the obvious way, and in place of Equation (36) we obtain:

$$p_1 \exp\left(-\int_0^x \mu^1_{02}(t)dt\right) + p_2 \exp\left(-\int_0^x \mu^2_{02}(t)dt\right) = \frac{(1-p)^{-1} - e^{\Lambda(x)}}{(1-p)^{-1}e^{\Lambda(x)}}. \tag{37}$$

We remark that the common life-table assumption of an upper limit ω to lifetimes implies that $\mu(x)$ explodes at $x = \omega$. In ordinary survival analysis this is unproblematic, because $\int_0^x \mu(t)dt$ need not be bounded, but here the bound on $\Lambda(x)$ may present a fundamental limit to inference if penetrance is high and data are sparse.

4.4 *Allowing for Genetic Testing*

Genetic testing is easily included by adding transitions into 'tested' states to the model in Figure 3, but we use a slightly different approach which is easier to extend later. Figure 5 shows a Markov model representing the information gained from a test. At birth, the Mendelian probability of having a mutation is $p < 1$, but after testing it is either 0 or 1. The state space of this model is $\mathcal{P} = \{1, 2\}$, and the state occupied at age x by the i^{th} life is the process $\mathbf{P}^i(x)$. Combining the two models, we work with the state space $\mathcal{S} \times \mathcal{P}$, and the state occupied at age x is the process $(\mathbf{S}^i(x), \mathbf{P}^i(x))$. Our counting processes are now $\mathbf{N}^i_{jk,l}(x)$ and $\mathbf{N}^i_{j,kl}(x)$ representing jumps from state (j, l) to (k, l), and from state (j, k) to (j, l), respectively, and the indicator of presence in state (j, k) is denoted $\mathbf{Y}^i_{j,k}(x)$.

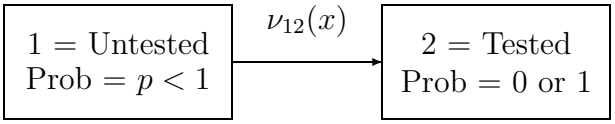

Figure 5: A Markov model of transfers between states representing the Mendelian probability (at birth) of having a mutation, and genetic testing.

All that we are interested in is that $\mathbf{P}^i(x)$ is observable; we are not interested in the intensity in Figure 5, even though it could conceivably depend on $\mu_{02}(x)$. The processes we can observe are $\mathbf{N}_1^i(x) = \mathbf{N}_{02,1}^i(x) + \mathbf{N}_{12,1}^i(x)$ and $\mathbf{N}_{02,2}^i(x)$. Define $\mathbf{Y}_1^i(x) = \mathbf{Y}_{0,1}^i(x) + \mathbf{Y}_{1,1}^i(x)$, and let G^i be the natural filtration generated by the observable processes. Assuming $\mu_{12}(x) \approx 0$, $\mathbf{N}_1^i(x)$ has G^i-compensator $\int_0^x \mathbf{Y}_1{}^i(t)\lambda(t,\mu)\mu_{02}(t)dt$, while $\mathbf{N}_{02,2}^i(x)$ has G^i-compensator $\int_0^x \mathbf{Y}_{0,2}^i(t)\mu_{02}(t)dt$.

Sum over all M lives, dropping the i superscripts, and define $\mathbf{J}_1(x)$ and $\mathbf{J}_{0,2}(x)$ in the same way as before. Then:

$$\hat{\Lambda}(x) = \int_0^x \frac{\mathbf{J}_1(t)}{\mathbf{Y}_1(t)} d\mathbf{N}_1(t) + \int_0^x \frac{\mathbf{J}_{0,2}(t)}{\mathbf{Y}_{0,2}(t)} d\mathbf{N}_{02,2}(t) \tag{38}$$

is a Nelson-Aalen-type estimate of $\int_0^x (1 + \lambda(t,\mu))\mu_{02}(t)dt$, or:

$$\hat{\Gamma}(x) = \int_0^x \frac{\mathbf{J}_1(t)}{\mathbf{Y}_1(t)\lambda(t,\mu)} d\mathbf{N}_1(t) + \int_0^x \frac{\mathbf{J}_{0,2}(t)}{\mathbf{Y}_{0,2}(t)} d\mathbf{N}_{02,2}(t) \tag{39}$$

is a Nelson-Aalen-type estimate of $2\int_0^x \mu_{02}(t)dt$, and we can proceed as before.

By orthogonality of the (compensated) components of a multivariate counting process:

$$\mathrm{Var}[\hat{\Lambda}(x)] \approx \int_0^x \frac{\mathbf{J}_1(t)(\mathbf{Y}_1(t) - \Delta\mathbf{N}_1(t))}{(\mathbf{Y}_1(t))^3} d\mathbf{N}_1(t) + \int_0^x \frac{\mathbf{J}_{0,2}(t)(\mathbf{Y}_{0,2}(t) - \Delta\mathbf{N}_{02,2}(t))}{(\mathbf{Y}_{0,2}(t))^3} d\mathbf{N}_{02,2}(t) \tag{40}$$

(we omit $\mathrm{Var}[\hat{\Gamma}(x)]$). For confidence intervals of $\mu_{02}(x)$, we can use the Weird Bootstrap, simulating onset of EOAD separately among those who have or have not been tested.

In practice, if either risk group is rather small, it might be better to omit it, because Equation (39) suggests that any cases of onset in that group will be extremely influential.

4.5 *Allowing for Unaffected Relatives*

Figure 6 extends the model of Figure 5 to allow for the subject to be born before the carrier status of the at-risk parent or any siblings is known (State 0). (This was called risk group 4 in Section 2.2.) As soon as the parent or any sibling is known to have a mutation, the subject's mutation probability at birth is fixed, usually at $p = 0.5$ (State 1). Alternatively, the subject could be tested first. The state space now is $\mathcal{S} \times \mathcal{P}$, where $\mathcal{P} = \{0, 1, 2\}$.

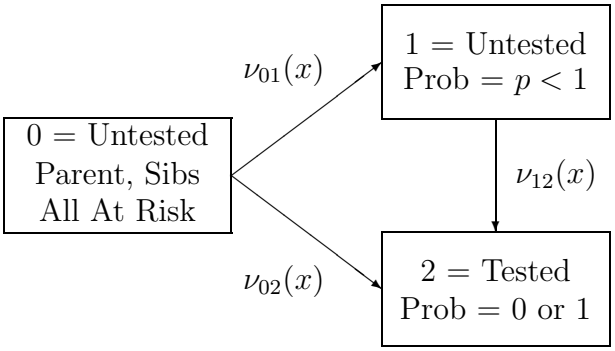

Figure 6: A Markov model of transfers between states representing probability at birth of having a mutation. For the parent and siblings to be 'at risk' means it is not known whether or not any of them have a mutation. Transition from State 0 to State 1 represents learning that the parent or a sibling does have a mutation; this fixes p (usually at 1/2).

Note that $\nu_{01}(x)$ will be a function of $\mu_{02}(t)$ (in general, at all ages $t \leq x$) since that transition represents onset of EOAD in a relative. However, we need not try to estimate $\nu_{01}(x)$; all that matters is that the transition is observable.

Having no affected relatives is the least tractable part of the model. In previous sections, the observable counting process components, in respect of the i^{th} life, had compensators $\int_0^x \mathbf{Y}_1{}^i(t)\lambda(t,\mu)\mu_{02}(t)dt$ and $\int_0^x \mathbf{Y}^i_{0,2}(t)\mu_{02}(t)dt$ in which the integrands took the general form:

$$\text{Indicator} \ \times \ \text{Function of age and } \mu_{02}(t). \tag{41}$$

The key point is that the second term was the same for all lives in the risk group; the compensator depending on the circumstances of the i^{th} life only through the indicator. This is no longer true if the subject has only unaffected parents and siblings, or in general no affected relatives in any number of generations. The compensator can still be put in the form of Equation (41), but now the second term will in general be a different function, say $\lambda^i(\mu)$, for each life, depending on the relatives' ages as well. As an estimate we get (in the obvious notation):

$$\hat{\Lambda}(x) = \int_0^x \frac{\mathbf{J}_1(t)}{\mathbf{Y}_1(t)} d\mathbf{N}_1(t) + \int_0^x \frac{\mathbf{J}_{0,2}(t)}{\mathbf{Y}_{0,2}(t)} d\mathbf{N}_{02,2}(t) + \sum_{i=1}^{i=M} \int_0^x \frac{\mathbf{J}^i_0(t)}{\mathbf{Y}^i_0(t)} d\mathbf{N}^i_0(t) \tag{42}$$

(we omit $\hat{\Gamma}(x)$) in which the third term is unhelpful, and the Nelson-Aalen methodology breaks down. The simplest solution is to exclude time spent in State 0 of Figure 6, and use the estimate of Section 4.4, at the cost of not using all of the data. This will not lead to bias: the compensators leading to Equation (38) are not changed by anything in this section, so the estimate has the same statistical properties as before. The form of Newcombe's (1981) estimate did allow this information to be used.

4.6 *Remarks*

(a) The fundamental difference between this approach and the life table method of Newcombe (1981) is in the conditioning. We condition (probabilities of) events that may befall a person age x on information known at age x, not on information acquired later, whereas Newcombe (1981) conditions on the last known risk status at the time of the investigation. Some statistical properties of the Nelson-Aalen-type estimates are available, while those of the life table estimates appear not to be.

(b) This method requires that the time spent in different risk groups can be observed or approximated; therefore, in future, the ages at which genetic tests were taken will be a relevant part of the pedigree.

(c) It might be thought that, in time, complete families will have genetic tests, so that non-carriers can be excluded without bias. However, the uptake of genetic tests is quite low, in respect of severe untreatable disorders, so this is only likely to happen once effective treatments are available. Epidemiologists might have to deal with mixtures of risk groups for some time to come.

5. Estimates of Incidence Rates

5.1 *Choice of Estimator*

In practically all of the families studied, some genetic tests have been carried out, suggesting Equations (38) or (39) as estimates. No information is given on when tests were taken, but we can assume that all tests were very recent, so almost no time has been spent in the tested state. Therefore, we can drop the second terms in Equation (38) or Equation (39). Should genetic testing become more widespread, the second term in Equation (38) or Equation (39) will begin to contribute.

5.2 *Approximations Used With Pedigrees*

First, we have to discard any incomplete sibships in pedigrees, since only complete sibships can be regarded as random samples of mutated and wild-type alleles, based on the appropriate Mendelian probabilities. Most published pedigrees give complete sibships in respect of more recent generations but not older generations.

We need two items of information, in respect of each life included; unfortunately neither is always straightforward. They are:

(a) The age at which each person entered the risk group (entered State (0,1) or (1,1) in the model, which being unknown) which is their age when their parent or first sibling contracted EOAD. (In fact, we found only one example in which a sibling contracted EOAD before the parent (who had died young) but this was excluded for other reasons.) This, however, is not usually known from the pedigree, because we do not know the parent's age when each of their children was born. This datum may well exist in the full pedigree, but it is not usually published.

(b) The age at onset of EOAD, or the age at censoring in other cases. The problem here is that the ages at censoring or death of unaffected siblings are often not given, even when complete sibships are shown.

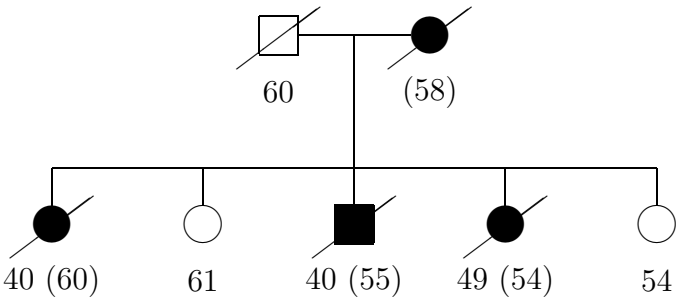

Figure 7: Pedigree of the family with the Leu282Arg PSEN-1 mutation (Aldudo *et al.*, 1998), see Appendix A. Squares are males, circles are females, and a slash denotes death. Affected individuals are shown as filled squares/circles. The age at onset or oldest observed age free of AD is shown, and the age at death is given in brackets. By convention siblings are listed left-to-right in birth order.

Here, we explain the approximations we have used, making all possible use of the information that is in the pedigrees. We shall see that the effect of (a) above should be small, but that (b) is more serious.

First, we assume that the affected parent's age at the birth of his or her children is 30 years, on average. Given the parent's age at onset, this allows us to approximate the ages at which their children entered the $p = 1/2$ risk group. For example, Figure 7 shows a pedigree of a family with the Leu282Arg PSEN-1 mutation (Aldudo *et al.*, 1998), see Appendix A. The average age of onset in this family is 43 years (Aldudo *et al.*, 1998). If the average age at childbirth is 30 years, then the average age of a child when their parent suffers onset is 13 years. We use this whenever the exact age is unknown, or cannot be better approximated. We proceed as follows:

(a) The mother is excluded since her age at onset is unknown.
(b) We assume the first, third and fourth children were 13 when the mother suffered onset.
(c) We can do slightly better with the two surviving children. Their average age is about 58 years, so we suppose their mother suffered onset $58 - (43 - 30) = 45$ years ago. Then the second child was 16 and the fifth child 9 at that time.

Despite some sweeping assumptions, this approach is not likely to affect the estimation significantly. The reason is that ages at onset are roughly in the range 30–60 years. Unless child-bearing extends very far on either side of age 30, there should be few cases in which the true age on entering the relevant risk group is within this range. Therefore:

(a) most errors will occur at ages with no, or few, observed cases of onset (so it might be unsuitable for use with diseases that occur before age 20, say); and
(b) any errors will tend to occur at ages where the exposures are greatest, and therefore their effect will be minimal.

Next, if the ages of unaffected siblings are not known, we use the convention that they are listed in birth order to estimate highest and lowest ages at which censoring might have

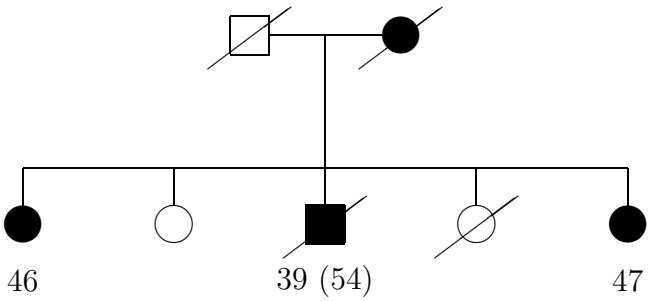

Figure 8: A hypothetical sibship in which the ages of unaffected lives are not known. The oldest sibling is known to be age 59.

occurred. Estimates based on these extremes then define a feasible region for $\hat{\mu}_{02}(x)$. For example, consider the hypothetical sibship shown in Figure 8. There are two unaffected sisters, one still alive.

(a) The older must be at least 55 (say), since her younger brother died at age 54, but she cannot be older than about 58, since her older sister is alive at age 59.

(b) The younger could have died in infancy, so may never have entered the $p = 1/2$ risk group. At the other extreme, she could be just about two years younger than her unaffected sister, and have died recently. So the highest and lowest possible ages at censoring are about 56 and 0 respectively.

In a few cases, extra information (such as the age at onset of an affected child of a sibling) helps to refine these bounds, and we are not interested in the exact age at censoring if it is above 60 anyway. We had to exclude many sibships for which even these crude calculations could not be made. Figures 9, 10 and 11 show the approximate maximum and minimum exposure times in the $p = 1/2$ risk group, for males and females combined and separately. Note that the combined samples include some siblings whose sex was not identified in the pedigrees.

The need to use approximate bounds on the exposures is a significant weakness of this analysis, but one that could only be overcome with access to original pedigree data. We suggest that this is an important area for consideration in future research.

5.3 *Smoothing Method*

We have six samples: males and females combined and separately, each with minimum and maximum estimated exposures in the $p = 1/2$ risk group. We estimate $\hat{\Lambda}(x)$ for each, then smooth them using a biweight-kernel method, and Equation (32) is then solved numerically. (Smoothing is also helpful because the ages of onset show clustering at some quinquennial ages, suggesting approximations in the data.) Kernel smoothing is natural since it exploits the stochastic integral framework of Nelson-Aalen estimators, and it is quite robust to the choice of kernel. See Andersen *et al.* (1993) for details, including choice of bandwidth and treatment of extreme ages.

A kernel-smoothed estimate $\tilde{\Lambda}(x)$ with bandwidth b is a weighted average of $\hat{\Lambda}(y)$ for $y \in [x - b, x + b]$. The weights are provided by the kernel, which is just a symmetric

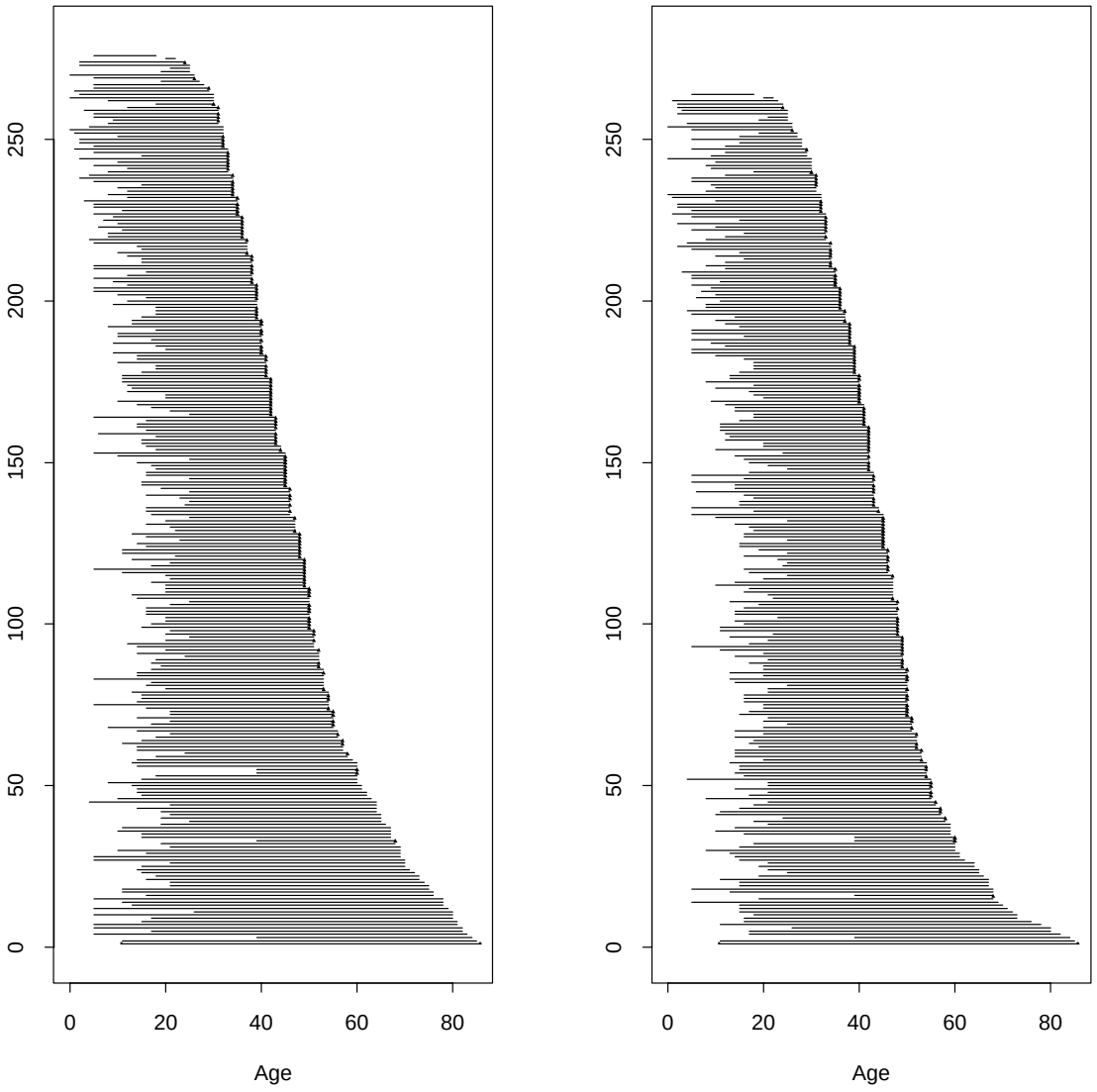

Figure 9: Estimated exposure times for all persons in the $p = 1/2$ risk group. Each line represents the time spent in the risk group by a single individual. The estimated maximum exposure times are on the left (276 lives), and minimum exposure times on the right (264 lives). Exposures ending with onset of EOAD are indicated by a triangle.

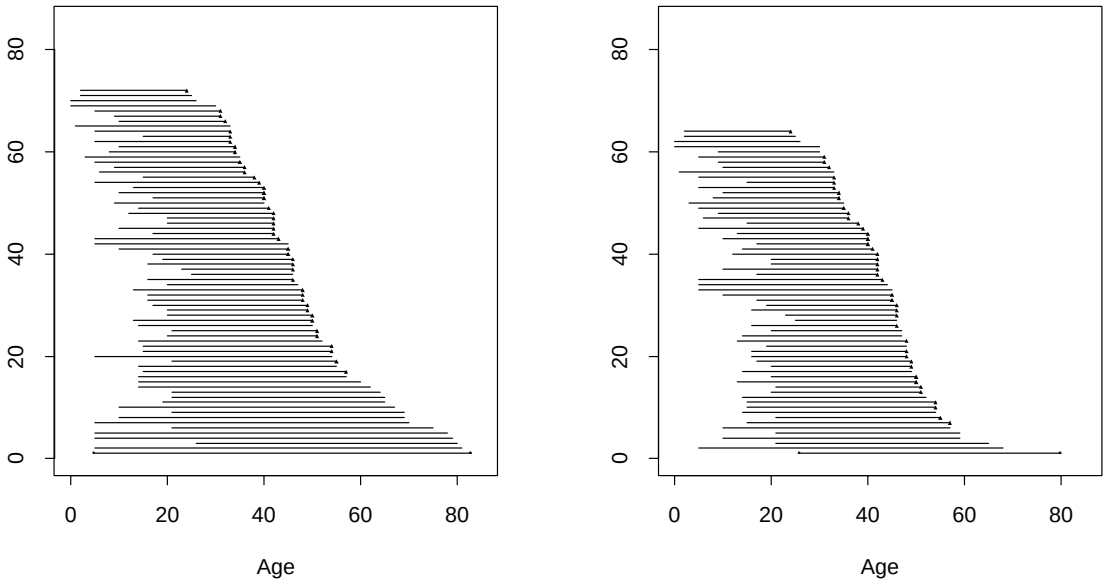

Figure 10: Estimated exposure times for men, maximum exposure times are on the left (72 lives), and minimum exposure times on the right (64 lives). Exposures ending with onset of EOAD are indicated by a triangle.

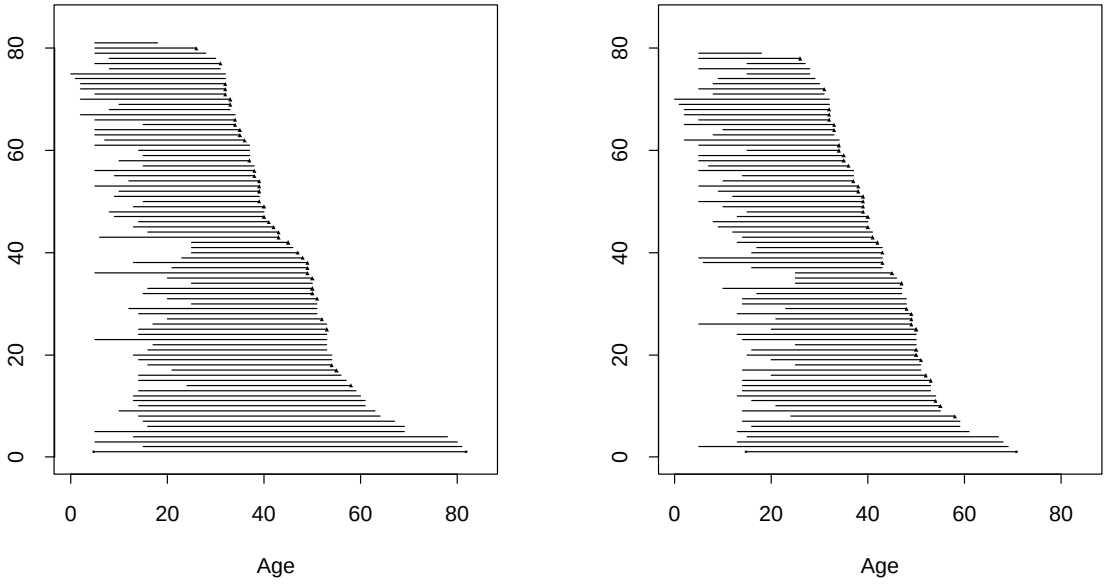

Figure 11: Estimated exposure times for women, maximum exposure times are on the left (81 lives), and minimum exposure times on the right (79 lives). Exposures ending with onset of EOAD are indicated by a triangle.

function on $[-b, b]$ integrating to 1. The biweight kernel is given on $[-1, 1]$ by:

$$K(t) = \frac{15}{16}(1 - t^2)^2 \tag{43}$$

and it is zero outside $[-1, 1]$ It gives progressively heavier weight to points close to 0 in the interval $[-1, 1]$. We scale the kernel by the bandwidth b, so that:

$$\tilde{\Lambda}(x) = \frac{1}{b} \int_{-\infty}^{\infty} K\left(\frac{x-u}{b}\right) \hat{\Lambda}(u) du. \tag{44}$$

Let x_L and x_U be the extreme ages. For $x_L < x < x_L + b$, we use an asymmetric kernel, K_q(t). Letting $q = x/b$, then on $[-1, q]$ we have $K_q(t) = K(t)(\alpha + \beta t)$, where

$$\alpha = \frac{64(8 - 24q + 48q^2 - 45q^3 + 15q^4)}{(1+q)^5(81 - 168q + 126q^2 - 40q^3 + 5q^4)} \tag{45}$$

and:

$$\beta = \frac{1120(1-q)^3}{(1+q)^5(81 - 168q + 126q^2 - 40q^3 + 5q^4)}. \tag{46}$$

For $x_U - b < x < x_U$, we let q = $(x_U - x)/b$ and replace t with $-t$ in K_q(t). Note that we are not interested in ages over about 60, since later onset is probably not EOAD, but the data include cases of onset at higher ages; therefore the adjustment at the upper age limit does not affect the results reported here.

The bandwidth b is chosen to minimise the mean integrated squared error (MISE) of $\lambda(x, \mu)\mu_{02}(x)$ over a suitable range τ_L to τ_U (we use $\tau_U < x_U$ since we are not interested in ages over about 60). For convenience write $\lambda(x, \mu)\mu_{02}(x) = \alpha(x)$. We have:

$$\text{MISE}(b) = \text{E}\left[\int_{\tau_L}^{\tau_U} [\hat{\alpha}(u) - \alpha(u)]^2 du\right]. \tag{47}$$

Klein & Moeschberger (1997) show that it suffices to minimise

$$G(b) = \sum_{i=1}^{J-1} (\frac{u_{i+1} - u_i}{2})[\tilde{\alpha}^2(u_i) + \tilde{\alpha}^2(u_{i+1})] - 2b^{-1}\left[\sum_{i \neq j} K(\frac{t_i - t_j}{b})\Delta\hat{\Lambda}(t_i)\Delta\hat{\Lambda}(t_j)\right] \tag{48}$$

where the first sum is evaluated at J suitable points $\tau_L = u_1 < u_2 < ... < u_J = \tau_U$, and the second sum at the jump-times of $\hat{\Lambda}(x)$. Figure 12 shows $G(b)$ for the six samples, and Table 1 shows the results.

Figure 13 shows the the unsmoothed $\hat{\Lambda}(x)$ and smoothed $\tilde{\Lambda}(x)$ for men and women, combined and separately, using the maximum and minimum estimated exposures for each. In all cases, $\hat{\Lambda}(x)$ exceeds its theoretical bound of $\log 2$ (Section 4.3) by about age 50, so we shall be unable to estimate $\mu_{02}(x)$ beyond that age.

Table 1: Optimal bandwidths for biweight kernel smoothing.

	Maximum Exposures			Minimum Exposures		
Sex	No. of Lives	τ_U	Optimal Bandwidth	No. of Lives	τ_U	Optimal Bandwidth
Combined	276	60	4.5 years	264	60	4.5 years
Male	72	56	2.3 years	64	56	2.2 years
Female	81	56	2.3 years	79	56	2.3 years

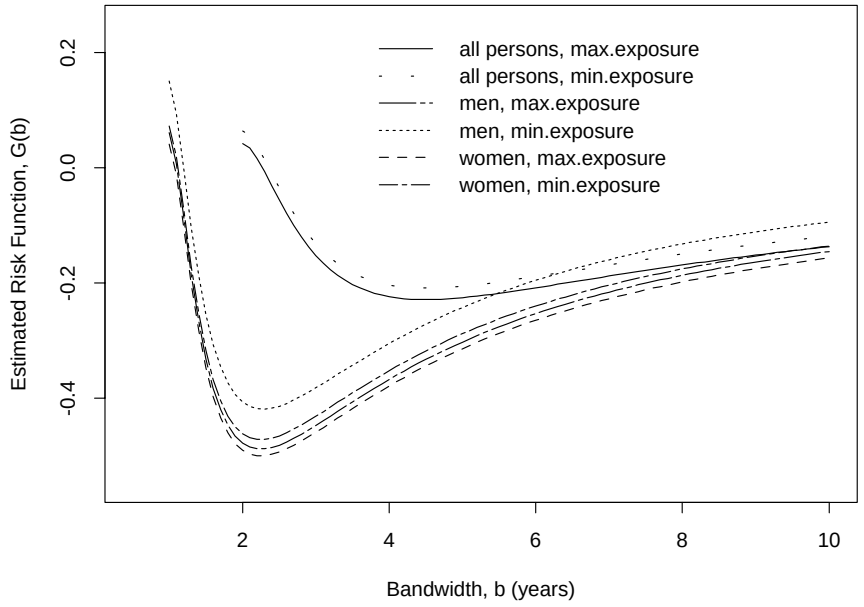

Figure 12: Estimated risk function, G(b), for use in determining the optimal bandwidth for the PSEN-1 mutation data.

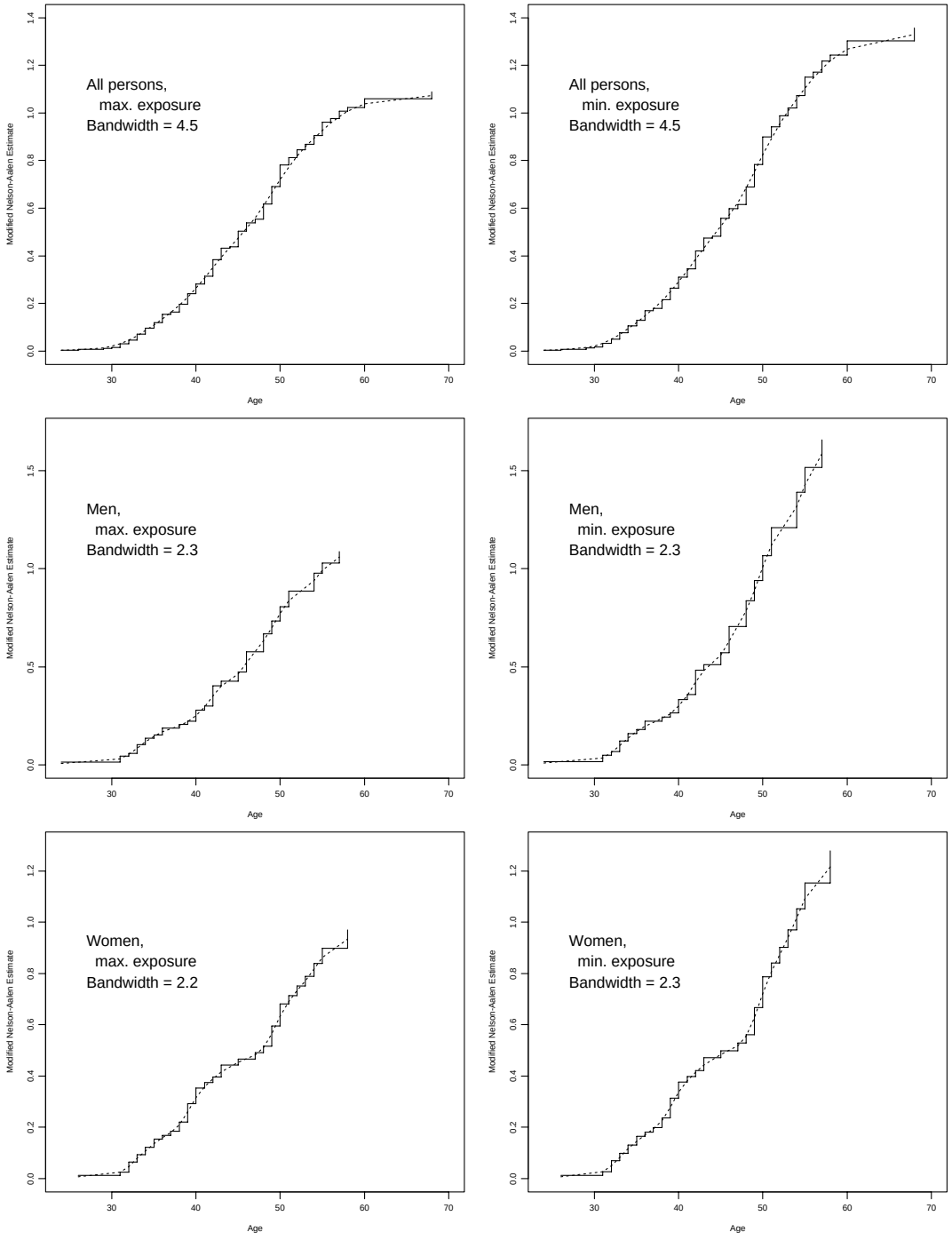

Figure 13: Modified Nelson-Aalen estimates of $\int_0^x \lambda(t,\mu)\mu_{02}(t)dt$ for all persons, men and women, in families with PSEN-1 mutations. The smoothed versions are shown as dashed lines. Note different scales.

5.4 *Bootstrap Confidence Intervals*

For approximate confidence limits, we generate 500 random samples of onset based on the observed exposures and onsets at each age (the 'Weird Bootstrap', see Section 4.1). For each of these samples, the intensity is calculated as before, and at each age the 25^{th} and 475^{th} samples give an approximate 95% confidence interval for $\hat{\mu}_x^{02}$.

5.5 *Results*

Figure 14 shows the resulting estimates of $\mu_{02}(x)$. This has several features:

(a) Although estimates are obtained up to about age 50 (when $\hat{\Lambda}(x)$ exceeds $\log 2$) their behaviour changes at about age 45.
(b) The confidence limits are limited to shorter age ranges than the estimates, because in each case, among the 500 simulated experiences there were some in which $\hat{\Lambda}(x)$ exceeded $\log 2$ at a lower age than in the actual sample.
(c) The estimates in respect of males and females show some unevenness that may be evidence of clustering at certain ages, but those for the combined sample seem to be well-smoothed.
(d) The general features of all the estimates are the same; unevenness of the smaller samples aside, $\mu_{02}(x)$ reaches about 0.1 – 0.2 by age 45.

For practical use, it would probably be best not to use these intensities directly, but to use them as a guide in choosing some simpler, smoother function that may be extrapolated to age 60 (say).

Probabilities of survival free of EOAD ($\exp(-\int_0^x \mu_{02}(t)dt)$) are shown in Figure 15, with bootstrapped 95% confidence limits. What is perhaps most significant is that in all cases, these survival probabilities are very low (< 0.5) by about age 45, so our inability to obtain good estimates beyond that age may be of little practical significance. That is, of course, a consequence of the high penetrance of PSEN-1 mutations.

6. Conclusions

6.1 *Estimation*

Genetic tests have created new risk groups in families that carry dominant single-gene disorders. We have extended Newcombe's (1981) life table method of estimating rates of onset to allow for these, in a Nelson-Aalen framework. Such methods will be needed to analyse pedigrees in which genetic testing is incomplete, which is likely to be the case in the absence of effective treatments.

6.2 *Limits on Estimation*

In the case of a single risk group with Mendelian probability p at birth of carrying a mutation, the function $\Lambda(x)$ being estimated has a finite upper bound, that may present an intrinsic limit to inference from relatively small samples (in this model framework).

6.3 *Information in Pedigrees*

We have identified the following data that should be included in pedigrees, in respect of each complete sibship, to allow the time spent in various risk groups to be found:

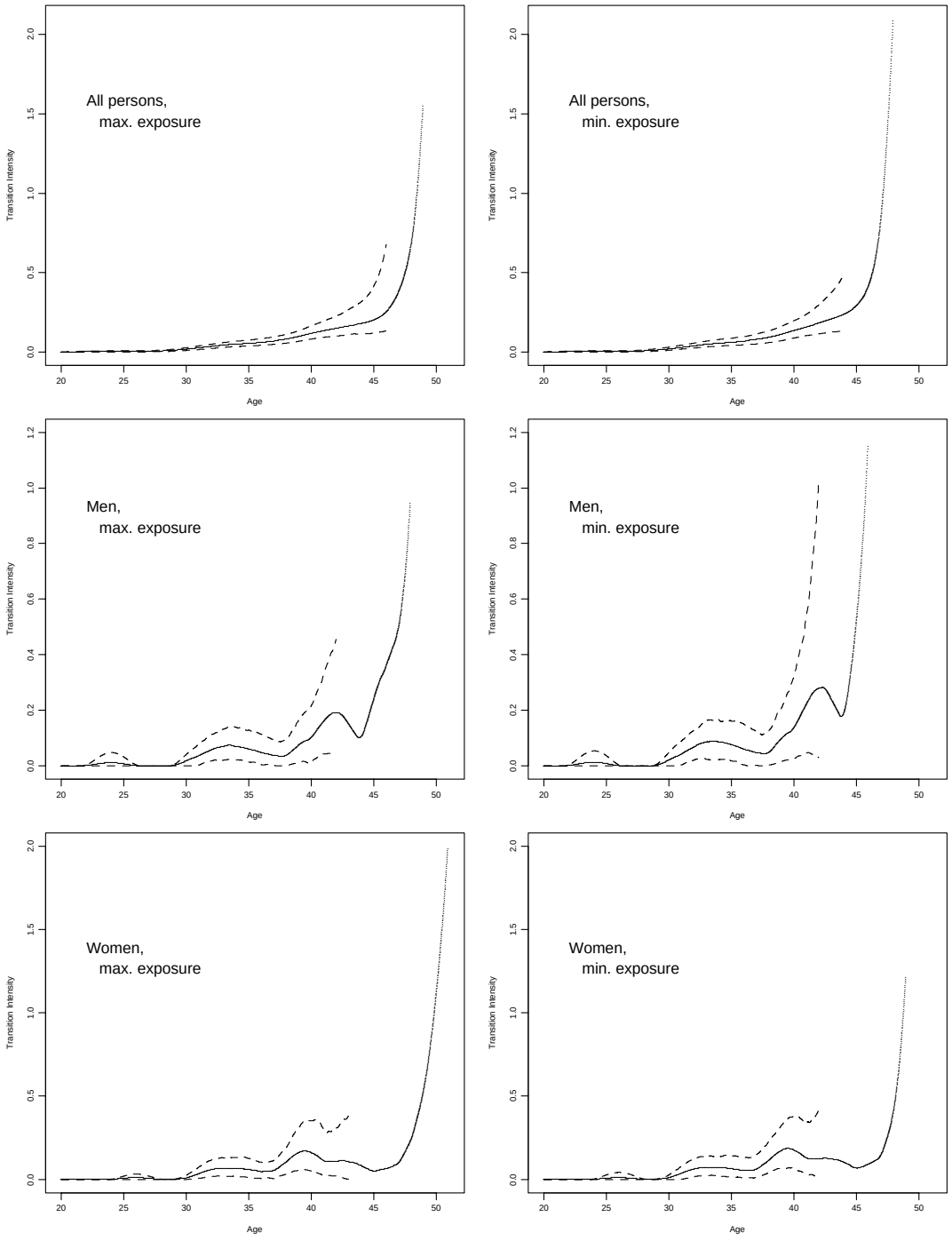

Figure 14: Estimated incidence rates of EOAD among men and women with PSEN-1 mutations, combined and separately, with approximate 95% confidence limits. Note different scales.

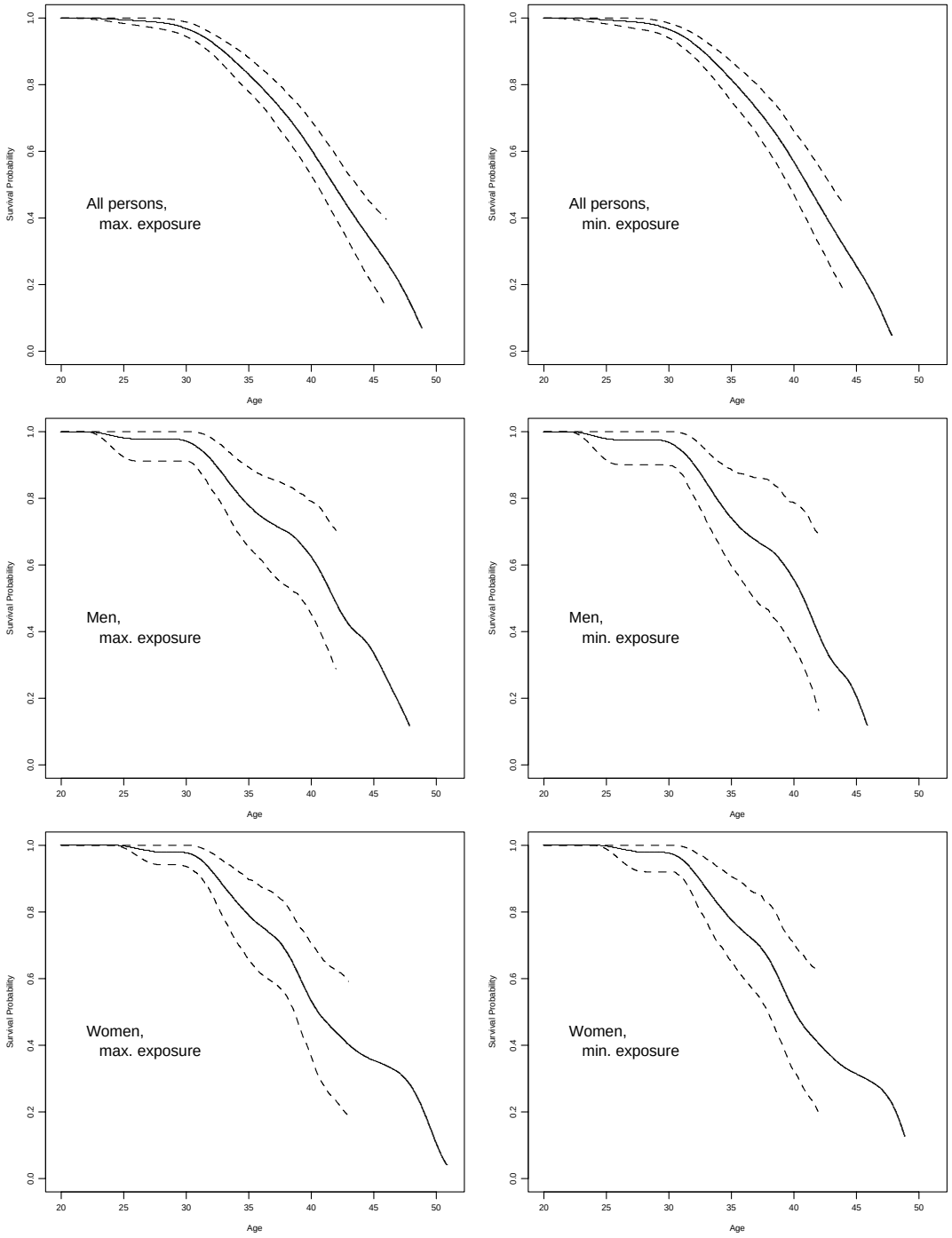

Figure 15: Probabilities of surviving free of EOAD among persons with PSEN-1 mutations, with approximate 95% confidence limits.

(a) the ages at which any genetic tests were taken;
(b) the parents' ages at the birth of each child; and
(c) the age at censoring or death of all unaffected siblings.

At present, genetic testing is so recent that (a) above is not yet important, but it will be necessary in future.

6.4 *Presenilin-1 Gene Mutations*

We have estimated rates of onset of EOAD based on published pedigrees of PSEN-1 mutations. While we were hampered by missing data, chiefly in respect of unaffected siblings of sufferers, we were able to obtain upper and lower estimates based on approximate minimum and maximum exposure times respectively. These showed enough consistency to be useful, especially since they were sufficiently high (roughly 0.1 – 0.2 by age 45) as to demonstrate penetrance exceeding 50%. If more complete pedigrees were available, it would be possible to extend the estimates to all relevant ages.

We have not tried to model differences in ages at onset among different families, but that is a worthwhile subject for future research.

6.5 *Applications*

This paper provides intensities that may be used to study the effect of genetic tests for PSEN-1 mutations on insurance, both for individuals (in terms of possible premium loadings if insurers do use genetic test information) and for other policyholders (possible costs of adverse selection if insurers do not use genetic test information). That will be the topic of further work.

Acknowledgements

This work was carried out at the Genetics and Insurance Research Centre at Heriot-Watt University. We would like to thank the sponsors for funding, and members of the Steering Committee for helpful comments at various stages. We thank Marc Cruts for help in obtaining some data.

References

Aldudo, J., Bullido, M.J., Arbizu, T., Oliva, R. & Valdivieso, F. (1998). Identification of a novel mutation (Leu282Arg) of the human presenilin 1 gene in Alzheimer's disease. *Neuroscience Letters*, **240**, 174–176.

Aldudo, J., Bullido, M.J., Arbizu, T. & Valdivieso, F. (1999). DGGE method for the mutational analysis of the coding and proximal promoter regions of the Alzheimer's disease presenilin-1 gene: 2 novel mutations. *Human Mutation*, **14**, 433–439.

Andersen, P.K., Borgan, Ø., Gill, R.D. & Keiding, N. (1993). *Statistical models based on counting processes.* Springer-Verlag, New York.

Aoki M., Abe K., Oda N., Ikeda M., Tsuda T., Kanai M., Shoji M., St. George-Hyslop P.H. & Itoyama Y. (1997). A presenilin-1 mutation in a family with Alzheimer's disease and distinctive abnormalities on cranial MRI. *Neurology*, **48**, 1118-1120.

Axelman K., Basun H. & Lannfelt L. (1998). Wide range of onset in a family with Alzheimer's disease and a His163Tyr mutation in the presenilin-1 gene. *Archives of Neurology*, **55**, 698-702.

Besançon, R., Lorenzi, A., Cruts, M., Radawiec, S., Sturtz, F., Broussolle, E., Chazot, G., Van Broeckhoven, C., Chamba, G. & Vandenberghe, A. (1997). Missense mutation in exon 11 (codon 378) of the presenilin-1 gene in a French family with early-onset Alzheimer's disease and transmission study by mismatch enhanced allele specific amplification. *Human Mutation (Mutation in Brief)*, **141**, online.

Bird T.D., Levy-Lahad E., Poorkaj P., Sharma V., Nemens E., Lahad A., Lampe T.H. & Schellenberg G.D (1996). Wide range in age of onset for chromosome 1–related familial Alzheimer's disease. *Annals of Neurology*, **40(6)**, 932-936.

Boteva, K., Vitek, M., Mitsuda, H., de Siva, H., Xu, P.-T., Small, G. & Gilbert, J.R. (1996). Mutation analyisis of presenilin 1 gene in Alzheimer's disease. *Lancet*, **347**, 130–131.

Campion, D., Flaman, J.-M., Brice, A., Hannequin, D., Dubois, B., Martin, C., Moreau, V., Charbonnier, F., Didierjean, O., Tardieu, S., Penet, C., Puel, M., Pasquier, F., Le Doze, F., Bellis, G., Calenda, A., Heilig, R., Martinez, M., Mallet, J., Bellis, M., Clerget-Darpoux, F., Agid, Y. & Frebourg, T. (1995a). Mutations of the presenilin 1 gene in families with early-onset Alzheimer's disease. *Human Molecular Genetics*, **4(12)**, 2373–2377.

Campion, D., Brice, A., Hannequin, D., Tardieu, S., Dubois, B., Calenda, A., Brun E., Penet C., Tayot J., Martinez, M., Bellis, G., Mallet, J., Agid, Y. & Clerget-Darpoux, F. (1995b). A large pedigree with early-onset Alzheimer's disease: clinical, neuropathological, and genetic characterization. *Neurology*, **45**, 80–85.

Campion, D., Brice, A., Dumanchin, C., Puel, M., Baulac, M., De La Sayette, V., Hannequin, D., Duyckaerts, C., Michon, A., Martin, C., Moreau, V., Penet, C., Martinez, M., Clerget-Darpoux, F., Agid, Y. & Frebourg, T. (1996). A novel presenilin 1 mutation resulting in familial Alzheimer's disease with an onset age of 29 years. *NeuroReport*, **7**, 1582–1584.

Campion, D., Dumanchin, C., Hannequin, D., Dubois, B., Belliard, S., Puel, M., Thomas-Anterion, C., Michon, A., Martin, C., Charbonnier, F., Raux, G., Camuzat, A., Penet, C., Mesnage, V., Martinez, M., Clerget-Darpoux, F., Brice, A. & Frebourg, T. (1999). Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum. *American Journal of Human Genetics*, **65**, 664–670.

Cervenakova L., Brown P., Sandoval F., Foncin J-F., Garruto R., Polincky R.J., Nee L & Goldfarb L. (1996). Identification of presenilin-1 gene point mutations in early-onset Alzheimer's disease families. *American Journal of Human Genetics*, **59(Suppl)**, A252.

Chapman J., Asherov A., Wang N., Treves T.A., Korczyn A.D. & Goldfarb L.G. (1995). Familial Alzheimer's disease associated with S182 codon 286 mutation. *Lancet*, **346**, 1040.

Clark R.F., Hutton M., Fuldner R.A., Froelich S., Karran E., Talbot C., Crook R., Lendon C., Prihar G., He C., Korenblat K., Martinez A., Wragg M., Busfield F., Behrens M.I., Myers A., Norton J., Morris J., Mehta N., Pearson C., Lincoln C., Baker M., Duff K., Zehr C., Perez-Tur J., Houlden H., Ruiz

A., Ossa J., Lopera F., Arcos M., Madrigal L., Collinge J., Humphreys C., Ashworth A., Sarner S., Fox N., Harvey R., Kennedy A., Roques P., Cline R.T., Phillips C.A., Venter J.C., Forsell L., Axelman K., Lilius L., Johnston J., Cowburn R, Viitanen M., Winblad B., Kosik K., Haltia M., Poyhonen M., Dickson D., Mann D., Neary D., Snowdon J., Lantos P., Lannfelt L., Rossor M., Roberts G.W., Adams M.D., Hardy J. & Goate A. (Alzheimer's Disease Collaborative Group) (1995). The structure of the presenilin 1 (S182) gene and identification of six novel mutations in early onset AD families. *Nature Genetics*, **11**, 219–222.

Crook, R., Verkkoniemi, A., Perez-Tur, J., Mehta, N., Baker, M., Houlden, H., Farrer, M., Hutton, M., Lincoln, S., Hardy, J., Gwinn, K., Somer, M., Paetau, A., Kalimo, H., Ylikoski, R., Pöyhönen, M., Kucera, S. & Haltia, M. (1998). A variant of Alzheimer's disease with spastic paraparesis and unusual plaques due to deletion of exon 9 of presenilin 1. *Nature Medicine*, **4(4)**, 452–455.

Cruts, M., Backhovens, H., Wang, S.-Y., Van Gassen, G., Theuns, J., De Jonghe, C., Wehnert, A., De Voecht, J., de Winter, G., Cras, P., Bruyland, M., Datson, N., Weissenbach, J., den Dunnen, J.T., Martin, J.-J., Hendriks, L. & Van Broeckhoven, C. (1995). Molecular genetic analysis of familial early-onset Alzheimer's disease linked to chromosome 14q24.3. *Human Molecular Genetics*, **4(12)**, 2363–2371.

Cruts, M., Hendriks, L. & Van Broeckhoven, C. (1996). The presenilin genes: a new gene family involved in Alzheimer's disease. *Human Molecular Genetics*, **5 (Review)**, 1449–1455.

Cruts, M., van Duijn, C.M., Backhovens, H., Van den Broeck, M., Wehnert, A., Serneels, S., Sherrington, R., Hutton, M., Hardy, J., St. George-Hyslop, P.H., Hofman, A. & Van Broeckhoven, C. (1998). Estimation of the genetic contribution of presenilin-1 and -2 mutations in a population-based study of presenile Alzheimer disease.. *Human Molecular Genetics*, **7(1)**, 43–51.

Cruts, M. & Van Broeckhoven, C. (1998). Presenilin mutations in Alzheimer's disease. *Human Mutation*, **11**, 183–190.

Czech, C., Tremp, G. & Pradier, L. (1999). Presenilins and Alzheimer's disease: biological functions and pathogenic mechanisms. *Progress in Neurobiology*, **60**, 363–384.

Dartigues, J.F. & Letenneur, L. (2000). Genetic epidemiology of Alzheimer's disease. *Current Opinion in Neurology*, **13**, 385–389.

De Jonghe C., Cruts M., Rogaeva E.A., Tysoe C., Singleton A., Vanderstichele H., Meschino W., Dermaut B., Vanderhoeven I., Backhovens H., Vanmechelon E., Morris C.M., Hardy J., Rubinsztein D.C., St. George-Hyslop P.H. & Van Broeckhoven C. (1999). Aberrant splicing in the presenilin-1 intron 4 mutation causes presenile Alzheimer's disease by increased Abeta42 secretion. *Human Molecular Genetics*, **8**, 1539-1540.

Dumanchin, C., Brice, A., Campion, D., Hannequin, D., Martin, C., Moreau, V., Agid, Y., Martinez, M., Clerget-Darpoux, F., Frebourg, T. & the French Alzheimer's Disease Study Group (1998). De novo presenilin 1 mutations are rare in clinically sporadic, early onset Alzheimer's disease cases. *Journal of Medical Genetics*, **35**, 672–673.

Elandt-Johnson, R.C. (1973). Age-at-onset distribution in chronic diseases. A life table approach to analysis of family data. *Journal of Chronic Disability*, **26**, 529–545.

Ezquerra, M., Carnero, C., Blesa, R., Gelpí, J.L., Ballesta, F. & Oliva, R. (1999). A presenilin 1 mutation (Ser169Pro) associated with early-onset AD and myoclonic seizures. *Neurology*, **52**, 566–570.

Ezquerra, M., Carnero, C., Blesa, R. & Oliva, R. (2000). A novel presenilin 1 mutation (Leu166Arg) associated with early-onset Alzheimer's disease. *Archives of Neurology*, **57**, 485–488.

Farlow M.R., Murrell J.R., Hulette C.M. & Ghetti B. (2000). Hereditary lateral sclerosis and Alzheimer disease associated with mutation at codon 261 of the presenilin (PS1) gene. *Neurobiology of Aging*, **21(Suppl)**, S62.

Finckh, U., Müller-Thomsen, T., Mann, U., Eggers, C., Markseiner, J., Meins, W., Binetti, G., Alberici, A., Hock, C., Nitsch, R.M. & Gal, A. (2000). High prevalence of pathogenic mutations in patients with early-onset dementia detected by sequence analyses of four different genes. *American Journal of Human Genetics*, **66**, 110–117.

Fischer, E.-P. & Berberich, K. (1999). *Impact of modern genetics on insurance.* Publications of the Cologne Re, No. 42.

Fleming, T.R. & Harrington, D.P. (1991). *Counting processes and survival analysis.* John Wiley, New York.

Forsell, C., Froelich, S., Axelman, K., Vestling, M., Cowburn, R.F., Lilius, L., Johnston, J.A., Engvall, B., Johansson, K., Dahlkild, A., Ingelson, M., St. George-Hyslop, P.H. & Lannfelt, L. (1997). A novel pathogenic mutation (Leu262Phe) found in the presenilin 1 gene in early-onset Alzheimer's disease. *Neuroscience Letters*, **234**, 3–6.

Fox N.C., Kennedy A.M., Harvey R.J., Lantos P.L., Roques P.K., Collinge J., Hardy J., Hutton M., Stevens J.M., Warrington E.K. & Rossor M.N. (1997). Clinicopathological features of familial Alzheimer's disease associated with the M139V mutation in the presenilin 1 gene. Pedigree but not mutation specific age at onset provides evidence for a further genetic factor. *Brain*, **120**, 491-501.

Fraser, P.E., Yang D-S., Yu, G., Lévesque L., Nishimura M., Arawaka S., Serpell L.C., Rogaeva E. & St George-Hyslop P. (2000). Presenilin structure, function and role in Alzheimer disease. *Biochimica et Biophysica Acta*, **1502**, 1-15.

Funato, H., Enya, M., Yoshimura, M., Morishima-Kawashima, M. & Ihara, Y. (1999). Presence of sodium dodecyl sulfate-stable amyloid beta-protein dimers in the hippocampus CA1 not exhibiting neurofibrillary tangle formation. *American Journal of Patholology*, **155(1)**, 23–28.

Gómez-Isla, T., Wasco, W., Pettingell, W.P., Gurubhagavatula, S., Schmidt, S.D., Jondro, P.D., McNamara, M., Rodes, L.A., DiBlasi, T., Growdon, W.B., Seubert, P., Schenk, D., Growdon, J.H., Hyman, B.T. & Tanzi, R.E. (1997). A novel presenilin-1 mutation: increased β-amyloid and neurofibrillary changes. *Annals of Neurology*, **41**, 809–813.

Gustafson, L., Brun, A., Englund, E., Hagnell, O., Nilsson, K., Stensmyr, M., Ohlin, A.-K. & Abrahamson, M. (1998). A 50-year perspective of a family with chromosome-14-linked Alzheimer's disease. *Human Genetics*, **102(3)**, 253–257.

Hardy, J. (1997). Amyloid, the presenilins and Alzheimer's disease. *Trends in Neuroscience*, **20(4)**, 154–159.

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.

Harvey R.J., Ellison D., Hardy J., Hutton M., Roques P.K., Collinge J., Fox N.C. & Rossor M.N. (1998). Chromosome 14 familial Alzheimer's disease: the clinical and neuropathological characteristics of a family with a leucine → serine (L250S) substitution at codon 250 of the presenilin 1 gene. *Journal of Neurology, Neurosurgery & Psychiatry*, **64**, 44-49.

Higuchi S., Yoshino A., Matsui T., Matsushita S., Satoh A., Iimura T., Ishikawa M., Arai H & Shirakura K. (2000). A novel PS1 mutation (W165G) in a Japanese family with early-onset Alzheimer's disease. *Alzheimer's Reports*, **3(4)**, 227-231.

Hiltunen M., Helisalmi S., Mannermaa A., Alafuzoff I., Koivisto A.M., Lehtovirta M., Pirskanen M., Sulkava R., Verkkoniemi A & Soininen H (2000). Identification of a novel 4.6-kb genomic deletion in presenilin-1 gene which results in exclusion of exon 9 in a Finnish early onset Alzheimer's disease family: an Alu core sequence-stimulated recombination ?. *European Journal of Human Genetics*, **8**, 259-266.

Hüll M., Fiebich B.L., Dykierek P., Schmidtke K., Nitzsche E., Orszagh M., Deuschl G., Moser E., Schumacher M., Lucking C., Berger M. & Bauer J. (1998). Early-onset Alzheimer's disease due to mutations of the presenilin-1 gene on chromosome 14: a 7-year follow-up of a patient with a mutation at codon 139. *European Archives of Psychiatry and Clinical Neurosciences*, **248**, 123-129.

Hutton M., Busfield F., Wragg M., Crook R., Perez-Tur J., Clark R.F., Prihar G., Talbot C., Phillips H., Wright K., Baker M., Lendon C., Duff K., Martinez A., Houlden H., Nichols A., Karran E., Roberts G., Roques P., Rossor M., Venter J.C., Adams M.D., Cline R.T., Phillips C.A., Fuldner R.A., Hardy J. & Goate A. (1996). Complete analysis of the presenilin 1 gene in early onset Alzheimer's disease. *Neuroreport*, **7**, 801-805.

Ikeda, M., Sharma, V., Mark Sumi, S., Rogaeva, E.A., Poorkaj, P., Sherrington, R., Nee, L., Tsuda, T., Oda, N., Watanabe, M., Aoki, M., Shoji, M., Abe, K., Itoyama, Y., Hirai, S., Schellenberg, G.D., Bird, T.D. & St. George-Hyslop, P.H. (1996). The clinical phenotype of two missense mutations in the presenilin 1 gene in Japanese patients. *Annals of Neurology*, **40**, 912–917.

Jacquier M., Arango D., Torres O., Cruts M., Serrano M., Matallana M., Cano C. & Van Broeckhoven C. (2000). Presenilin mutations in a Colombian familial and sporadic sample. *Neurobiology of Aging*, **21(Suppl)**, S176.

Jørgensen, P., Bus, C., Pallisgaard, N., Bryder, M. & Jørgensen, A.L. (1996). Familial Alzheimer's disease co-segregates with a Met146Ile substitution in presenilin-1. *Clinical Genetics*, **50**, 281–286.

Kamimura, K., Tanahashi, H., Yamanaka, H., Takahashi, K., Asada, T. & Tabira, T. (1998). Familial Alzheimer's disease genes in Japanese. *Journal of the Neurological Sciences*, **160**, 76–81.

Kamino, K., Sato, S., Sakaki, Y., Yoshiiwa, A., Nishiwaki, Y., Takeda, M., Tanabe, H., Nishimura, T., Ii, K., St. George-Hyslop, P.H., Miki, T. & Ogihara, T. (1996). Three mutations of presenilin 1 gene in early-onset Alzheimer's disease families. *Neuroscience Letters*, **208**, 195–198.

Klein, J.P. & Moeschberger, M.L. (1997). *Survival analysis: Techniques for censored and truncated data.* Springer-Verlag, New York.

Kowalska A., Forsell C., Florczak J., Pruchnik-Wolinska D., Modestowicz R., Paprzycki W., Wender M. & Lannfelt L. (1999). A Polish pedigree with Alzheimer's disease determined by a novel mutation in exon 12 of the presenilin 1 gene: clinical and molecular characterization. *Folia Neuropathologica*, **37(1)**, 415-424.

Kwok, J.B.J., Taddei, K., Hallupp, M., Fisher, C., Brooks, W.S., Broe, G.A., Hardy, J., Fulham, M.J., Nicholson, G.A., Stell, R., St. George-Hyslop, P.H., Fraser, P.E., Kakulas, B., Clarnette, R., Relkin, N., Gandy, S.E., Schofield, P.R. & Martins, R.N. (1997). Two novel (M233T and R278T) presenilin-1 mutations in early-onset Alzheimer's disease pedigrees and preliminary evidence for association of presenilin-1 mutations with a novel phenotype. *NeuroReport*, **8**, 1537–1542.

Kwok, J.B.J., Smith M.J., Brooks W.S., Kril J., Mclean C., Hallupp M. & Schofield P.R. (2000). Variable presentation of Alzheimer's disease and/or spastic paraparesis phenotypes in pedigrees with a novel PS-1 exon 9 gene deletion or exon 9 splice acceptor mutations. *Neurobiology of Aging*, **21(Suppl)**, S25.

Lendon C.L., Martinez A., Behrens I.M., Kosik K.S., Madrigal L., Norton J., Neuman R., Myers A., Busfield F., Wragg M., Arcos M., Arango Viana J.C., Ossa J., Ruiz A., Goate A.M. & Lopera F. (1997). E280A PS-1 mutation causes Alzheimer's disease but age of onset is not modified by ApoE alleles. *Human Mutation*, **10**, 186-195.

Macdonald, A.S. (1996). An actuarial survey of statistical models for decrement and transition data III: Counting process models. *British Actuarial Journal*, **2**, 703–726.

Macdonald, A.S. (1997). How will improved forecasts of individual lifetimes affect underwriting?. *Philosophical Transactions of the Royal Society* B, **352**, 1067–1075, and (with discussion) *British Actuarial Journal*, **3**, 1009–1025 and 1044–1058.

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

Macdonald, A.S. & Pritchard, D.J. (2000). A mathematical model of Alzheimer's disease and the ApoE gene. *ASTIN Bulletin*, **30**, 69–110.

Macdonald, A.S. & Pritchard, D.J. (2001). Genetics, Alzheimer's disease and long-term care insurance. *To appear in North American Actuarial Journal.*

Martin, J.J., Gheuens, J., Bruyland, M., Cras, P., Vandenberghe, A., Masters, C.L., Beyreuther, K., Dom, R., Ceuterick, C., Lübke, U., Van Heuverswijn, H., De Winter, G. & Van Broeckhoven, C. (1991). Early-onset Alzheimer's disease in 2 large Belgian families. *Neurology*, **41**, 62–68.

Morelli L., Prat M.I., Levy E., Mangone C.A. & Castano E.M. (1998). Presenilin 1 Met146Leu variant due to an A → T transversion in an early-onset familial Alzheimer's disease pedigree from Argentina. *Clinical Genetics*, **53**, 469-473.

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

Palmer, M.S., Beck, J.A., Campbell, T.A., Humphries, C.B., Rogues, P.K., Fox, N.C., Harvey, R., Rossor, M.N. & Collinge, J. (1999). Pathogenic presenilin 1 mutations (P436S & I143F) in early-onset Alzheimer's disease in the U.K.. *Human Mutation*, **13(3)**, 256.

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

Perez-Tur, J., Froelich, S., Prihar, G., Crook, R., Baker, M., Duff, K., Wragg, M., Busfield, F., Lendon, C., Clark, R.F., Roques, P., Fuldner, R.A., Johnston, J., Cowburn, R., Forsell, C., Axelman, K., Lilius, L., Houlden, H., Karran, E., Roberts, G.W., Rossor, M., Adams, M.D., Hardy, J., Goate, A., Lannfelt, L. & Hutton, M. (1995). A mutation in Alzheimer's disease destroying a splice acceptor site in the presenilin-1 gene. *NeuroReport*, **7**, 297–301.

Perez-Tur, J., Croxton, R., Wright, K., Phillips, H., Zehr, C., Crook, R., Hutton, M., Hardy, J., Karran, E., Roberts, G.W., Lancaster, S. & Haltia, T. (1996). A further presenilin 1 mutation in the exon 8 cluster in familial Alzheimer's disease. *Neurodegeneration*, **5**, 207–212.

Poduslo S.E., Herring K. & Neal M (1996). A presenilin 1 mutation in an early onset Alzheimer's family: no association with presenilin 2. *Neuroreport*, **7**, 2018-2020.

Poorkaj P., Sharma V., Anderson L., Nemens E., Alonso M.E., Orr H., White J., Heston L., Bird T.D. & Schellenberg G.D. (1998). Missense mutations in the chromosome 14 familial Alzheimer's disease presenilin 1 gene. *Human Mutation*, **11**, 216-221.

Prihar G., Verkkoniem A., Perez-Tur J., Crook R., Lincoln S., Houlden H., Somer M., Paetau A., Kalimo H., Grover A., Myllykangas L., Hutton M., Hardy J & Haltia M (1999). Alzheimer's disease PS-1 exon 9 deletion defined. *Nature Medicine*, **5**, 1090.

Queralt R., Ezquerra M., Castellvi M., Lleo A., Blesa R. & Oliva R. (2001). Detection of the presenilin 1 gene mutation (M139T) in early-onset familial Alzheimer's disease in Spain. *Neuroscience Letters*, **299**, 239-241.

Ramirez-Dueñas, M.G., Rogaeva, E.A., Leal, C.A., Lin, C., Ramirez-Casillas, G.A., Hernandez-Romo, J.A., St. George-Hyslop, P.H. & Cantu, J.M. (1998). A novel Leu171Pro mutation in presenilin-1 gene in a Mexican family with early-onset Alzheimer disease. *Annals Génét*, **41(3)**, 149–153.

Raux G., Gantier R., Thomas-Anterion C., Boulliat J., Verpillat P., Hannequin D., Brice A., Frebourg T. & Campion D. (2000a). Dementia with prominent frontotemporal features associated with L113P presenilin 1 mutation. *Neurology*, **55**, 1577-1579.

Raux G., Gantier R., Martin C., Pothin Y., Brice A., Frebourg T. & Campion D. (2000b). A novel presenilin 1 missense mutation (L153V) segregating with early-onset autosomal dominant Alzheimer's disease. *Human Mutation*, **16**, 95.

Reznik-Wolf, H., Treves, T.A., Davidson, M., Aharon-Peretz, J., St. George-Hyslop, P.H., Chapman, J., Korczyn, A.D., Goldman, B. & Friedman, E. (1996a). A novel mutation of presenilin 1 in familial Alzheimer's disease in Israel detected by denaturing gradient gel electrophoresis. *Human Genetics*, **98**, 700–702.

Reznik-Wolf, H., Machado J., Haroutunlan V., DeMarco L., Walter G.F., Chapman Y., Goldman B., Davidson M., Johnston J., Lannfelt L., Dani S.U. & Friedman E. (1996b). Somatic mutational anaylsis of the APP and presenilin 1 genes in Alzheimer's disease patients brains using denaturing gradient gel electrophoresis (DGGE). *Neurobiology of Aging*, **17(Suppl)**, S14.

Rogaev, E.I., Sherrington, R., Rogaeva, E.A., Levesque, G., Ikeda, M., Liang, Y., Chi, H., Lin, C., Holman, K., Tsuda, T., Mar, L., Sorbi, S., Nacmias, B., Piacentini, S., Amaducci, L., Chumakov, I., Cohen, D., Lannfelt, L., Fraser,

P.E., Rommens, J.M. & St. George-Hyslop, P.H. (1995). Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature*, **376**, 775–778.

Romero, I., Jørgensen, P., Bolwig, G., Fraser, P.E., Rogaeva, E., Mann, D., Havsager, A.-M. & Jørgensen, A.L. (1999). A presenilin-1 Thr116Asn substitution in a family with early-onset Alzheimer's disease. *NeuroReport*, **10**, 2255–2260.

Rossor, M.N., Fox, N.C., Beck, J., Campbell, T.C. & Collinge, J. (1996). Incomplete penetrance of familial Alzheimer's disease in a pedigree with a novel presenilin-1 gene mutation. *The Lancet*, **347**, 1560.

Russo, C., Schettini, G., Saido, T.C., Hulette, C., Lippa, C., Lannselt, L., Ghetti, B., Ganbetti, P., Tabaton, M. & Teller, J.K. (2000). Presenilin-1 mutations in Alzheimer's disease. *Nature*, **405**, 531–532.

St. George-Hyslop, P.H. (1998). Role of genetics in tests of genotype, status, and disease progression in early-onset Alzheimer's disease. *Neurobiology of Aging*, **19(2)**, 133–137.

St. George-Hyslop, P.H. (2000). Molecular genetics of Alzheimer's disease. *Biological Psychiatry*, **47**, 183–199.

Sandbrink, R., Zhang, D., Schaeffer, S., Masters, C.L., Bauer, J., Förstl, H. & Beyreuther, K. (1996). Missense mutations of the PS-1/S182 gene in German early-onset Alzheimer's disease patients. *Annals of Neurology*, **40(2)**, 265–266.

Sato, S., Kamino, K., Miki, T., Doi, A., Li, K., St. George-Hyslop, P.H., Ogihara, T. & Sakaki, Y. (1998). Splicing mutation of presenilin-1 gene for early-onset familial Alzheimer's disease. *Human Mutation*, **Suppl**, 591–594.

Sherrington, R., Rogaev, E.I., Liang, Y., Rogaeva, E.A., Levesque, G., Ikeda, M., Chi, H., Lin, C., Li, G., Holman, K., Tsuda, T., Mar, L., Foncin, J.-F., Bruni, A.C., Montesi, M.P., Sorbi, S., Rainero, I., Pinessi, L., Nee, L., Chumakov, I., Pollen, D., Brookes, A., Sanseau, P., Polinsky, R.J., Wasco, W., Da Silva, H.A.R., Haines, J.L., Pericak-Vance, M.A., Tanzi, R.E., Roses, A.D., Fraser, P.E., Rommens, J.M. & St. George-Hyslop, P.H. (1995). Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature*, **375**, 754–760.

Smith, M.J., Gardner, R.J.M., Knight, M.A., Forrest, S.M., Beyreuther, K., Storey, E., McLean, C.A., Cotton, R.G.H., Cappal, R., Masters, C.L. (1999). Early-onset Alzheimer's disease caused by a novel mutation at codon 219 of the presenilin-1 gene. *NeuroReport*, **10**, 503–507.

Smith M.J., Kwok J.B.J., Mclean C.A., Kril J.J., Broe G.A., Nicholson G.A., Cappai R., Hallupp M., Cotton R.G.H., Masters C.L., Schofield P.R. & Brooks W.S. (2001). Variable phenotype of Alzheimer's disease with spastic paraparesis. *Annals of Neurology*, **49**, 125-129.

Sorbi S., Nacmias B., Forleo P., Piacentini S., Sherrington R., Rogaev E, St. George-Hyslop P.H. & Amaducci L. (1995). Missense mutation of S182 gene in Italian families with early-onset Alzheimer's disease. *Lancet*, **346**, 439-440.

Steiner H., Revesz T., Neumann M., Romig H., Grim M.G., Pesold B., Kretzschmar H.A., Hardy J., Holton J.L., Baumeister R., Houlden H & Haass C. (2001). A pathogenic presenilin-1 deletion causes abberant A-beta-42 production in the absence of congophilic amyloid plaques. *Journal of Bioliogical Chemistry*, **276(46)**, 7233-7239.

Strachan, T. & Read, A.P. (1999). *Human molecular genetics, second edition.* BIOS Scientific Publishers, Oxford.

Subramanian, K., Lemaire, J., Hershey, J.C., Pauly, M.V., Armstrong, K. & Asch, D.A. (2000). Estimating adverse selection costs from genetic testing for breast and ovarian cancer: The case of life insurance. *Journal of Risk and Insurance*, **66**, 531–550.

Sugiyama, N., Suzuki, K., Matsumura, T., Kawanishi, C., Onishi, H., Yamada, Y., Iseki, E. & Kosaka, K. (1999). A novel missense mutation (G209P) in exon 8 of the presenilin 1 gene in a Japanese family with presenile familial Alzheimer's disease. *Human Mutation*, **14(1)**, 90–93.

Taddei, K., Kwok, J.B.J., Kril, J.J., Halliday, G.M., Creasey, H., Hallupp, M., Fisher, C., Brooks, W.S., Chung, C., Andrews, C., Masters, C.L., Schofield, P.R. & Martins, R.N. (1998). Two novel presenilin-1 mutations (Ser169Leu and Pro436Gln) associated with very early onset Alzheimer's disease. *NeuroReport*, **9**, 3335–3339.

Tanahashi H., Mitsunaga Y., Takahashi K., Tasaki H., Watanabe S. & Tabira T. (1995). Missense mutation of S182 gene in Japanese familial Alzheimer's disease. *Lancet*, **346**, 440.

Tanahashi H., Kawakatsu S., Kaneko M., Yamanaka H., Takahashi K. & Tabira T. (1996). Sequence analysis of presenilin-1 gene mutation in Japanese Alzheimer's disease patients. *Neuroscience Letters*, **218**, 139-141.

Tedde A., Forleo P., Nacmias B., Piccini C., Bracco L., Piacentini S & Sorbi S. (2000). A presenilin-1 mutation (Leu392Pro) in a familial AD kindred with psychiatric symptoms at onset. *Neurology*, **55**, 1590-1591.

Terreni L., Valeria C., Calella A.M., Gavazzi A., Alberoni M., Grimaldi L.M., Mariani C. & Forloni G. (2000). A novel missense mutation (L219F) in exon 8 of the presenilin 1 gene in an Italian family with presenile familial Alzheimer's disease. *Neurobiology of Aging*, **21(Suppl)**, S176-S177.

Tysoe C., Whittaker J., Xuereb J., Cairns N.J., Cruts M., Van Broeckhoven C., Wilcock G & Rubinsztein D.C. (1998). A presenilin-1 truncating mutation is present in teo cases with autopsy-confirmed early-onset Alzheimer's disease. *American Journal of Human Genetics*, **62**, 70-76.

Verkkoniemi A., Somer M., Rinne J.O., Myllykangas L., Crook R., Hardy J., Vitanen M., Kalimo H. & Haltia M. (2000). Variant Alzheimer's disease with spastic paraparesis: clinical characterization. *Neurology*, **54**, 1103-1109.

Warren, V., Brett, P., Macdonald, A.S., Plumb, R.H. & Read, A.P. (1999). *Genetic tests and future need for long-term care in the UK: report of a Work Group of the Continuing Care Conference Genetic Tests and Long-term Care Study Group.* Continuing Care Conference, London.

Wasco, W., Pettingell, W.P., Jondro, P.D., Schmidt, S.D., Gurubhagavatula, S., Rodes, L., Diblasi, T., Romano, D.M., Guenette, S.Y., Kovacs, D.M., Growdon, J.H. & Tanzi, R.E. (1995). Familial Alzheimer's chromosome 14 mutations. *Nature Medicine*, **1(9)**, 848.

Wisniewski, T., Dowjat, W.K., Buxbaum, J.D., Khorkova, O., Efthimiopoulos, S., Kulczycki, J., Lojkowska, W., Weigiel, J., Wisniewski, H.M. & Frangione, B. (1998). A novel Polish presenilin-1 mutation (P117L) is associated with familial Alzheimer's disease and leads to death as early as the age of 28 years. *NeuroReport*, **9**, 217–221.

Yasuda, M., Maeda, K., Ikejiri, Y., Kawamata, T., Kuroda, S. & Tanaka, C. (1997). A novel missense mutation in the presenilin-1 gene in a familial Alzheimer's disease pedigree with abundant amyloid angiopathy. *Neuroscience Letters*, **232**, 29–32.

Yasuda, M., Maeda, K., Hashimoto, M., Yamashita, H., Ikejiri, Y., Bird, T.D., Tanaka, C. & Schellenberg, G.D. (1999). A pedigree with a novel presenilin mutation at a residue that is not conserved in presenilin 2. *Archives of Neurology*, **56**, 65–69.

Yasuda, M., Maeda, S., Kawamata, T., Tamaoka, A., Yamamoto, Y., Kuroda, S., Maeda, K. & Tanaka, C. (2000). Novel presenilin-1 mutation with widespread cortical amyloid deposition but limited cerebral amyloid angiopathy. *Journal of Neurology, Neurosurgery and Psychiatry*, **68**, 220–223.

Younkin, S., Scheuner, D., Song, X., Eckman, C., Citron, M., Suzuki, N., Bird, T.D., Hardy, J., Hutton, M., Lannfelt, L., Levy-Lahad, E., Peskind, E., Poorkaj, P., Schellenberg, G., Tanzi, R., Viitanen, M., Wasco, W. & Selkoe, D. (1996). The presenilin 1 and 2 mutations linked to familial Alzheimer's disease increase the extracelluar concentration of amyloid β protein (Aβ) ending at Aβ42(43). *Neurobiology of Aging*, **17**, S38 Abstract.

APPENDIX A

PRESENILIN-1 MUTATIONS

The following table summarises the family studies in which PSEN-1 mutations have been found.

(a) The mutations are specified in the standard notation in the genetics literature: for example Ala79Val means that at the 79^{th} position in the amino acid chain, the amino acid Alanine has been replaced by Valine. Note that this refers to the product of transcription; the amino acid chain may be subject to further processing resulting in a shorter functional protein. The mutations are ordered by position.

(b) The pedigree IDs identify the particular family referred to in the study, with some information about ethnicity (when available). The following abbreviations have been used: Amer = American, Amer-C = American-Caucasian, Arg = Argentinian, Ashk-J = Ashkenazi-Jewish, Aus = Australian, Bel = Belgian, Brit = British, C = Caucasian, Col = Colombian, Dan = Danish, Eng = English, Fin = Finnish, Fr = French, Fr-Can = French-Canadian, G = German, It = Italian, J = Japanese, Mex = Mexican, Mex-A = Mexican-American, Pol = Polish, Rom-J = Romanian-Jewish, Russ-J = Russian-Jewish, Scot = Scottish, Scot-I = Scottish-Irish, Sp = Spanish, Swe = Swedish.

(c) The 'Used' column specifies whether the pedigree has been used for the estimates made in this paper.

(d) The key to the pedigree summary is as follows:

- c: Number of cases of EOAD in the family
- g: Number of generations affected
- a: Ages at onset of affected individuals
- m: Mean age at onset of affected individuals (n = No. included in mean)
- r: Range of ages at onset (minimum – maximum) of affected individuals

No.	Mutation	Pedigree ID	Used	Pedigree Summary	Reference
1	Ala**79**Val	1005, C	No	c:8, g:3, m:53	Cruts *et al.* (1998)
		1087, C	No	c:3, g:3, m:55	Cruts *et al.* (1998)
		1061, C	No	c:2, g:2, m:58	Cruts *et al.* (1998)
			No	m:64	St.George-Hyslop (1998)
		G	Yes	c:2, g:2, a:55, 58	Finckh *et al.* (2000)
2	Val**82**Leu	SAl 508, Fr	No	c:3, g:3, m:55, r:53–58	Campion *et al.* (1995a, 1999)
3	del**ΔI83/M84**	Scot	Yes	c:5, g:3, median:36	Steiner *et al.* (2001)
4	Val**94**Met	Col	No	a:53 (sporadic?)	Jacquier *et al.* (2000)
5	Val**96**Phe	OS-3. J	Yes	c:4, g:2, m: 52.5±5.07, r:49–60	Kamino *et al.* (1996)
6	Phe**105**Leu	G	Yes	c:3, g:3, a:50, 52, < 60	Finckh *et al.* (2000)
7	Leu**113**Pro	SAL 513, Fr	Yes	c:6, g:4, m:42.4±5.0, r:38–50	Raux *et al.* (2000a)
8	del **Intron4**	FD177, Eng	No	c:6, g:3	Tysoe *et al.* (1998), DeJonghe *et al.* (1999)
		142, Eng	No	c:1 (no family history)	Tysoe *et al.* (1998), De Jonghe *et al.* (1999)
		F105/160, Brit	No	c:17, g:6, m:37±3, r:36-40	De Jonghe *et al.* (1999)
		Tor122, Brit	No	c:10, g:3, m:37	De Jonghe *et al.* (1999)
		79/95, Brit	No	c:4, g:3, m:37	De Jonghe *et al.* (1999)
		593, Brit	No	c:1 (no family history)	De Jonghe *et al.* (1999)
9	Tyr**115**His	ALZ 025, Fr	No	c:2, g:2, a:35, 37	Campion *et al.* (1995a)
		ALZ 076, Fr	No	c:3, g:3, r:36–47	Campion *et al.* (1999)
10	Tyr**115**Cys	1066, C	No	c:10, g:4, m:45	Cruts *et al.* (1998), Hardy (1997)

No.	Mutation	Pedigree ID	Used	Pedigree Summary	Reference
11	Tyr**116**Asn	Dan	Yes	c:4, g:3, m:38, r:35–41	Romero *et al.* (1999)
12	Pro**117**Leu	Pol	Yes	c:8, g:3, m: 30.3±4.2, r:24–33	Wisniewski *et al.* (1998)
13	Glu**120**Asp	Rom-J	Yes	c:4, g:3, r:43–48	Reznik-Wolf *et al.* (1996a, 1996b), St.George-Hyslop (1998)
		ALZ 057, Fr	No	c:4, g:3, r:42–53	Campion *et al.* (1999)
		V, Brit	Yes	c:10, g:4, m:46.4±4.2, r:41-53	Poorkaj *et al.* (1998), Bird *et al.* (1996)
14	Glu**120**Lys	F121, Brit	No	c:6, g:3, r:32–39	Hutton*et al.* (1996), Hardy(1997)
15	Glu**123**Lys	ABCD-2, J	No	c:4, g:3, a:56, 62 (2 unknown)	Yasuda *et al.* (1999)
16	Asn**135**Asp	Mex-A	Yes	c:9, g:5, r:34–38	Crook *et al.* (1998)
17	Met**139**Ile	1674, C	No	details not reported	Boteva *et al.* (1996)
18	Met**139**Lys	ALZ 034, Fr	No	c:1, g:1, a:37 (isolated case)	Dumanchin *et al.* (1998)
19	Met**139**Thr	CAE 010, Fr	No	c:2, g:2, a:48, 50	Campion *et al.* (1995a, 1999)
		Sp	No	c:2, g:1, a:47, 48	Queralt *et al.* (2001)
20	Met**139**Val	F148, Brit	Yes	c:7, g:3, m:44.3, r:42–48	Clark *et al.* (1995), Hutton *et al.* (1996), Fox *et al.* (1997), Palmer *et al.* (1999)
		F206, Brit	Yes	c:9, g:4, m:37.7, r:36–40	Clark *et al.* (1995), Hutton *et al.* (1996), Fox *et al.* (1997), Palmer *et al.* (1999)
		G	No	c:6, g:3, r:42–44	Hüll *et al.* (1998)
		G	Yes	c:3, g:3, a:32 (2 unknown)	Finckh *et al.* (2000)
		G	No	c:1, g:1, a:40 (isolated case)	Sandbrink *et al.* (1996)
21	Ile**143**Phe	A, Brit	No	c:3, g:2, a:55, 53, 57	Rossor *et al.* (1996), Palmer *et al.* (1999)
22	Ile**143**Thr	AD/A, Bel	Yes	c:43, g:6, m:35.1±4.8, r:26–45	Cruts *et al.* (1995), Martin *et al.* (1991)
23	Met**146**Ile-a	Dan	Yes	c:7, g:3, m:44.4, r:38–57	Jørgensen *et al.* (1996), Cervenakova *et al.* (1996)
24	Met**146**Ile-b	E-M, Swe	Yes	c:6, g:4, m:42.7±6.3, r:35–49	Gustafson *et al.* (1998)
25	Met**146**Leu	FAD4(Okla1), It	No	c:8, g:3, m:45	Sherrington *et al.* (1995), Clark *et al.* (1995)
		Tor1.1, It	No	No details published	Sherrington *et al.* (1995)
		FAD4? It	No	c:3, m:45	Sorbi *et al.* (1995)
		It	No	c:3, m:36	Sorbi *et al.* (1995)
		It	No	m:35	Sorbi *et al.* (1995)
		It	No	3 from 3 families	Terreni *et al.* (2000)
		AR1, Arg	Yes	c:10, g:3, m:42±2.5 (gen II–III, n:6, r:40–46), m:35±2 (gen IV, n:4, r:33–38)	Morelli *et al.* (1998)
		ALZ 204, Fr	No	c:6, g:3, r:38–47	Campion *et al.* (1999)
26	Met**146**Val	Fin1	No	c:3, g:3; m:36	Clark *et al.*(1995)
		Man92/20	No	c:2, g:2, m:40	Clark *et al.*(1995)
		NY5201	No	c:7, g:2, m:37	Clark *et al.*(1995)
			No	c:1 (no other details)	Cervenakova *et al.* (1996)
27	Thr**147**Ile	ALZ 047, Fr	No	c:4, g:3, r:37–46	Campion *et al.* (1999)
28	Leu**153**Val	Fr	No	c:5, g:3, r:34–38	Raux *et al.* (2000b)
29	His**163**Arg	LH 603, Fr-Can	No	c:25, m:47.9±6.2, r:37–68	Bird *et al.* (1996), Sherrington *et al.* (1995), Boteva *et al.* (1996), Poorkaj *et al.* (1998)
		Tor 42, C	No	c:2, m:45	Sherrington *et al.* (1995)
		SAL 001, Fr	No	c:4, g:3, r:42–47	Campion *et al.* (1995a, 1999)
		TK-2, J	Yes	m:47.4±4.04, r:43–50	Kamino *et al.* (1996)
		H-1, J	No	c:14, m:39	Kamimura *et al.* (1998)
		J	No	c:1, a:38	Kamimura *et al.* (1998)
		J	No	c:1, a:51	Kamimura *et al.* (1998)
		Miy, J	No	c:2, m:45	Kamimura *et al.* (1998)
		HR1, G	Yes	c13, g:4, m:46.9±4.8, r:40–55	Bird *et al.* (1996), Poorkaj *et al.* (1998)

No.	Mutation	Pedigree ID	Used	Pedigree Summary	Reference
		J	No	c:14, g:5	Tanahashi *et al.* (1995)
		AD-Kae, J	No	c:1, a:41(sporadic)	Tanahashi *et al.* (1996)
		J	No	c:3, g:2, a: all late 40s	Poduslo *et al.* (1996)
		Fr	No	c:2 (no other details)	Cervenakova *et al.* (1996)
30	His**163**Ile	J	No	m:47	Kamino *et al.* (1996)
31	His**163**Tyr	Swed 2	No	c:8, g:3, m:47	Clark *et al.* (1995)
		Swed	Yes	c:22, g:4, m:54	Axelman *et al.* (1998)
32	Trp**165**Cys	ALZ 064, Fr	No	c:3, g:3, r:37–47	Campion *et al.* (1999)
33	Trp**165**Gly	J	Yes	c:5, g:3, m:35.8, r:34–38	Higuchi *et al.* (2000)
34	Leu**166**Arg	Sp	Yes	c:6, g:3, r:32–44	Ezquerra *et al.* (2000)
35	Ser**169**Leu	PERTH-4, Aus	Yes	c:3, g:2, a:31, 36, 39	Taddei *et al.* (1998)
36	Ser**169**Pro	Sp	Yes	c:4, g:2, a:33, 34, 35, 35	Ezquerra *et al.* (1999)
37	Leu**171**Pro	Ped 1, Mex	No	c:3, g:4, m:40	Ramirez-Dueñas *et al.* (1998)
		Ped 2-4, Mex	No	3 families, m:36, 37, 39	Ramirez-Dueñas *et al.* (1998)
38	Leu**173**Trp	ROU 118, Fr	No	c:2, g:2, a:24, 29	Campion *et al.* (1999)
39	Phe**177**Ser		No	details not published	Fraser *et al.* (2000)
40	Ser**178**Pro		No	details not published	Fraser *et al.* (2000)
41	Glu**184**Asp	ABCD-1, J	Yes	c:4, g:3, m:38.1±5.1	Yasuda *et al.* (1997)
42	Gly**206**Ser		No	details not published	Fraser *et al.* (2000)
43	Gly**209**Arg	J	Yes	c:3, g:2, m:49.6±3.1, a:46, 48, 53	Sugiyama *et al.* (1999)
44	Gly**209**Val	L, G	Yes	c:19, g:4, m:41.3±4.5, r:30–48	Bird *et al.* (1996), Cruts & Van Broeckhoven (1998), Poorkaj *et al.* (1998)
45	Ile**213**Thr	OS-2, J	Yes	c:4, g:2, m:45.0±4.24, r:42–48	Kamino *et al.* (1996)
46	Leu**219**Phe	It	No	details not published	Terreni *et al.* (2000)
47	Leu**219**Pro	MELB-1, Aus	Yes	c:10, g:5, a:47, 53, 54 (7 unknown)	Smith *et al.* (1999)
48	Gln**222**Ala		No	details not published	Fraser *et al.* (2000)
49	Ala**231**Thr	ALZ 043, Fr	No	c:4, g:3, m:52, r:45–57	Campion *et al.* (1995a, 1999)
50	Ala**231**Val	1072, C	No	c:6, g:3, m:58	Cruts *et al.* (1998), Hardy (1997)
51	Met**233**Leu	Sp	No	c:1, a:46 (isolated case)	Aldudo *et al.* (1999)
52	Met**233**Thr	PERTH-1, Aus	Yes	c:4, g:2, m:35	Kwok *et al.* (1997)
		ALZ 079, Fr	No	c:5, g:3, r:38–45	Campion *et al.* (1999)
53	Leu**235**Pro	SAL 510, Fr	Yes	c:4, g:2, m:32.5	Campion *et al.* (1996)
				c:5, g:3, r:29–39 (update)	Campion *et al.* (1999)
54	Ala**246**Glu	FAD 1, C	No	c:9, g:3, m:55	Sherrington *et al.* (1995)
55	Leu**250**Sel	F184, Brit	Yes	c:7, g:3, median:52, r:49–56	Hutton *et al.* (1996), Hardy(1997), Harvey *et al.* (1998)
56	Ala**260**Val	AM/JPN 1, J	No	c:9, g:3, m:40.3±5.5, r:27–46	Rogaev *et al.* (1995), Ikeda *et al.* (1996), Poorkaj *et al.* (1998)
57	Val**261**Phe		No	c:4, g:2, a:38 (3 unknown)	Farlow *et al.* (2000)
58	Leu**262**Phe	Swe	No	c:3, g:1, a:47, 48, 56	Forsell *et al.* (1997)
59	Cys**263**Arg	MGH12, C	No	c:5, m:50	Wasco *et al.* (1995)
			No	m:47	St.George-Hyslop (1998)
60	Pro**264**Leu	KG, Brit	Yes	c:5, g:2, m:43.2±1.6, r:41–45	Bird *et al.* (1996), Poorkaj *et al.* (1998)
		MGH6, C	No	c:2, g:1, a:45, 50	Wasco *et al.* (1995)
		SAL 511, Fr	No	c:6, g:3, r:45–56	Campion *et al.* (1995a, 1999)
		SAL 506, Fr	No	c:4, g:3, r:46–52	Campion *et al.* (1999)
		SAL 1633, Fr	No	c:4, g:3, r:51–55	Campion *et al.* (1999)
		EOFAD-6	No	m:39	Kwok *et al.* (1997)
61	Pro**267**Ser	F196, Brit	No	c:5, g:3, m:35, r:32–38	Clark *et al.* (1995), Hutton *et al.* (1996), Palmer *et al.* (1999)
62	Arg**269**Gly	Amer-C	No	c:3, g:2, a:47, 52, 54	Perez-Tur *et al.* (1996), Hardy (1997)
63	Arg**269**His	Amer-C	No	c:2, g:2, a:46, 61	Gómez-Isla *et al.* (1997)
		MAT-1, J	No	c:2, m:50	Kamimura *et al.* (1998), Hardy (1997)
64	Glu**273**Ala	Ok-1, J	No	c:2, m:63	Kamimura *et al.* (1998)
65	Arg**278**Thr	P-2, C	No	c:1, g:1, a:37 (isolated case)	Kwok *et al.* (1997)

No.	Mutation	Pedigree ID	Used	Pedigree Summary	Reference
66	Glu**280**Ala	C1, Col	No	c:11, g:3, m:49.9±6.9, r:41–55	Clark *et al.* (1995), Lendon *et al.* (1997)
		C2, Col	No	c:67, g:6, m:47.4±4.8, r:39–55	Clark *et al.* (1995), Lendon *et al.* (1997)
		C3, Col	No	c:10, g:4, m:52.0±4.9, r:41–59	Clark *et al.* (1995), Lendon *et al.* (1997)
		C4, Col	No	c:10 m:52.2±5.3, r:46–60	Lendon *et al.* (1997)
		C5, Col	No	c:9 m:47.4±11.9, r:36–62	Lendon *et al.* (1997)
		C8, Col	No	c:6 m:52.2±2.7, r:47–55	Lendon *et al.* (1997)
		C12, Col	No	c:1, a:42	Lendon *et al.* (1997)
		FAD-Ok, J	No	c:2, m:57	Tanahashi *et al.* (1996)
		F771	No	m:45	Clark *et al.* (1995)
		Col	No	c:2 (from separate families?)	Jacquier *et al.* (2000)
		COL	No	c:1, a:47	Kwok *et al.* (1997)
67	Glu**280**Gly	F168, Brit	No	c:5, g:3, m:41, r:39–42	Clark *et al.* (1995), Hutton *et al.* (1996)
		F183, Brit	No	c:2, g:3, m:43, r:42–45	Clark *et al.* (1995), Hutton *et al.* (1996)
		F196, Brit	No	c:1, g:1 (no age data)	Palmer *et al.* (1999)
68	Leu**282**Arg	Sp	Yes	m:43±5	Aldudo *et al.* (1998)
69	Ala**285**Val	SD-6, J	Yes	c:3, g:3, m:51	Ikeda *et al.* (1996)
		TOH-1, J	Yes	c:2, g:3, a:55, 45	Aoki *et al.* (1997)
		J	No	related to SD-6 or TOH-1 ?	Rogaev *et al.* (1995)
70	Leu**286**Val	FAD2	No	c:9, g:3, m:50	Sherrington *et al.* (1995)
		Ashk-J	No	c:4, g:2, a:42,47,48,50s	Chapman *et al.* (1995)
71	Ser**289**Cys		No	no age data	Cruts *et al.* (1996)
72	Ser**290**Cys		No	c:15, g:4, m:47.5±3.3	Sato *et al.* (1998), Hardy (1997)
73	**291–319** Del	AusAD-1 (EOFAD-3), Aus	Yes	c:13, g:3, m:45.8±6.1, r:36–54	Kwok *et al.* (1997, 1998, 2000), Smith *et al.* (2001)
		AusAD-2, Aus	No	c:4	Kwok *et al.* (2000)
		AusAD-3, Aus	No	No details published	Kwok *et al.* (2000)
		Finn2, Fin	Yes	c:22, g:4, m:50.9±5.2, r:40–61	Crook *et al.* (1998) Prihar *et al.* (1999), Verkkoniemi *et al.* (2000)
		Fin	Yes	c:4, g:2, a:43, 45, 42, 40	Hiltunen *et al.* (2000)
		F74, Brit	Yes	c:7, g:4, r:39–50	Perez-Tur *et al.* (1995), Hutton *et al.* (1996)
	(+Ser**290**Cys)	TK-1, J	No	c:15, g:4, m:47.5±3.3	Sato *et al.* (1998)
74	Glu**317**Gly		No	no age data	Hardy(1997)
75	Glu**318**Gly	ALZ 059, Fr	No	c:3, g:3, r:42–50	Campion *et al.* (1999)
		G	No	c:1, g:1, a:47 (isolated case)	Sandbrink *et al.* (1996)
		Swe	Yes	c:4, g:2, r:60–68	Forsell *et al.* (1997)
		PERTH-5, Aus	Yes	c:3, g:2	Taddei *et al.* (1998)
		1069	No	c:2, g:2, m:57	Cruts *et al.* (1998)
76	Gly**378**Glu	Fr	Yes	c:6, g:3, m:35, r:34–38	Besançon *et al.* (1997)
77	Gly**384**Ala	AD/B, Bel	No	c:25, g:5, m:34.7±3.0, r:30–39	Cruts *et al.* (1995), Martin *et al.* (1991)
		FAD-Yg, J	No	c:4, m:35	Tanahashi *et al.* (1996)
		Yg1, J	No	c:10, m:36	Kamimura *et al.* (1998)
78	Ser**390**Ile	ALZ 107, Fr	No	c:4, g:3, r:39–40	Campion *et al.* (1999)
79	Leu**392**Pro	It	Yes	c:4, g:3, m:38.3±4.0	Tedde *et al.* (2000)
80	Leu**392**Val	FAD R01, Fr	No	c:38, g:5, r:39–52	Campion *et al.* (1995b, 1999)
		It	No	results unpublished	Rogaev *et al.* (1995)
81	Asn**405**Ser	HI-1, J	No	c:1, g:1, a:48 (isolated case)	Yasuda *et al.* (2000)
82	Ala**409**Thr	Sp	No	c:1, a:58	Aldudo *et al.* (1999)
83	Cys**410**Tyr	FAD3(SNW), Russ-J	No	c:20, g:4?, m:51.7±2.7, r:48–56	Sherrington *et al.* (1995), Clark *et al.* (1995), Poorkaj *et al.* (1998)
		NIH2	No	No details published	Sherrington *et al.* (1995)
		ROU 011, Fr	No	c:14, g:4, r:40–60	Campion *et al.* (1995a, 1999)
			No	c:1 (no other details)	Cervenakova *et al.* (1996)

No.	Mutation	Pedigree ID	Used	Pedigree Summary	Reference
84	Leu**418**Phe		No	details not published	Fraser *et al.* (2000)
85	Leu**424**Arg		No	no details available	Kowalska *et al.* (1999)
86	Ala**426**Pro	HRX-III(XIII), Scot-I	Yes	c:6, g:2, m:46.0±3.5, r:41–51	Bird *et al.* (1996), Hardy (1997), Poorkaj *et al.* (1998)
87	Ala**431**Glu		No	details not published	Fraser *et al.* (2000)
88	Ala**434**Cys		No	details not published	Fraser *et al.* (2000)
89	Pro**436**Gln	SYD-1	Yes	c:2, g:2	Taddei (1998)
90	Pro**436**Ser	F223, Brit	Yes	c:2, g:2, a:44, 50	Palmer *et al.* (1999), Hardy (1997)
91	Ile**439**Val		No	details not published	Fraser *et al.* (2000)