A CORRECTION FOR ASCERTAINMENT BIAS IN ESTIMATING RATES OF ONSET OF HIGHLY PENETRANT GENETIC DISORDERS

BY

CAROLINA ESPINOSA AND ANGUS MACDONALD

ABSTRACT

Estimation of rates of onset of rare, late-onset dominantly inherited genetic disorders is complicated by: (a) probable ascertainment bias resulting from the 'recruitment' of strongly affected families into studies; and (b) inability to identify the true 'at risk' population of mutation carriers. To deal with the latter, Gui & Macdonald (2002a) proposed a non-parametric (Nelson-Aalen) estimate $\hat{\Lambda}(x)$ of a simple function $\Lambda(x)$ of the rate of onset at age x. The function $\Lambda(x)$ had a finite bound, which was an increasing function of the probability p that a child of an affected parent inherits the mutation and σ the life-time penetrance. However if $\hat{\Lambda}(x)$ exceeds this bound, it explodes to infinity, and this can happen at quite low ages. We show that such 'failure' may in fact be a useful measure of ascertainment bias. Gui & Macdonald assumed that $p = 1/2$ and $\sigma = 1$, but ascertainment bias means that $p > 1/2$ and $\sigma \neq 1$ in the sample. The maximum attained by $\hat{\Lambda}(x)$ allows us to estimate a range for the product $p\sigma$, and therefore the degree of ascertainment bias that may be present, leading to bias-corrected estimates of rates of onset. However, we find that even classical independent censoring, prior to ascertainment, can introduce new bias. We apply these results to early-onset Alzheimer's disease associated with mutations in the Presenilin-1 gene.

KEYWORDS

Ascertainment Bias; Early-Onset Alzheimer's Disease; Nelson-Aalen Estimate; Presenilin-1 Gene; Rate of Onset.

1. INTRODUCTION

1.1. Dominantly Inherited Gene Mutations

A few rare, but important, diseases of adulthood are caused by dominantly inherited mutations in a single gene; for example Huntington's disease (HD), early-onset Alzheimer's disease (EOAD) or familial breast cancer (BC). The age

Astin Bulletin 37(2), 429-452. doi: 10.2143/AST.37.2.2024075

at onset may be highly variable, often falling between 20 and 60 years, when people are working, raising families and possibly seeking insurance cover, so these disorders present social and ethical, as well as medical, problems.

In this paper we focus on EOAD. This can be caused by mutations in any one of three genes (more may yet be discovered) called Presenilin-1 (PSEN-1), Presenilin-2 (PSEN-2) and Amyloid Precursor Protein (APP). PSEN-1 mutations are highly likely to lead to EOAD, and studies of enough families have been published to allow rates of onset to be estimated.

A dominantly inherited disorder results from a mutation that triggers a deleterious process in some organ(s), so that inheriting just one copy of the mutated gene from either parent is sufficient to cause disease. If the process is of slow build-up (such as the accumulation of amyloid plaques in the brains of Alzheimer's disease (AD) sufferers) then disease onset may be delayed until well into adult life.

If the mutation is rare, we can ignore the small probability that both parents carry it, or that either parent carries two copies of the mutation (the latter might not be consistent with life anyway). Then the probability that any child of an affected parent will inherit the mutation, denoted p^*, is 1/2.

The variable age at onset is described by the penetrance function $q^*(x)$ defined as:

$$q^*(x) = P[\text{Mutation carrier suffers onset of disease not later than age } x] \quad (1)$$

assuming that all other decrements (including death) are absent, or equivalently the rate of onset $\mu^*(x)$ defined by:

$$q^*(x) = 1 - \exp\left(-\int_0^x \mu^*(t)\,dt\right). \quad (2)$$

Not all genetic disorders are fully penetrant, meaning that sometimes not all mutation carriers will develop the disorder. Define the lifetime penetrance σ^* as:

$$\sigma^* = \lim_{x \to \infty} q^*(x). \quad (3)$$

1.2. Problems of Ascertainment

Ideally, we could observe a population of mutation carriers and estimate $q^*(x)$ or $\mu^*(x)$ by ordinary survival analysis. However, unless everyone has had a fully reliable genetic test, mutation carriers are only discovered when they develop the disorder.

Alternatively, we could hope to observe a collection of complete generations of siblings, called sibships, sampled from the population of at-risk families in an unbiased manner. Then knowledge that $p^* = 1/2$ still allows us to estimate

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

TABLE 1

MODEL PARAMETERS WITH AND WITHOUT ASCERTAINMENT BIAS.

Parameter	Whole Population	Sampled Population
Probability of carrying a mutation	p^*	p
Lifetime penetrance	σ^*	σ
Penetrance function	$q^*(x)$	$q(x)$
Rate of onset	$\mu^*(x)$	$\mu(x)$

$q^*(x)$ or $\mu^*(x)$ by suitable conditioning in respect of persons not yet observed to have the disorder.

In practice, however, we may often obtain a sample of at-risk persons or sibships retrospectively, and in a highly non-random fashion (see Section 1.3). All the quantities defined above may then be different from their 'true' population values; we denote their new values by omitting the asterisks (see Table 1).

Note that the parameters p, σ and so on are not empirical estimates based on the actual sample, but the 'true' parameters in an expanded model, which now includes the mechanism by which sampling takes place. A major problem in genetic epidemiology, and the motivation for this paper, is that this sampling mechanism may be unknown. Indeed, in studies based on retrospective analysis of family histories collected from many sources (common in the study of rare disorders) it may not even be consistent within the sample.

1.3. Ascertainment Bias and Censored Data

In the case of rare genetic disorders, there is often some reason for a family coming to the attention of researchers (being ascertained), such as the number of affected members. We might expect large families, with multiple cases, to be overrepresented in many studies, leading to biased estimates of penetrance. It is a central problem in genetic epidemiology (see for example Sham (1998, Chapter 2), Thompson (1993)).

Adjusting for ascertainment bias requires a model of the mechanism for selecting families into a study, so ideally the data should be collected through a properly designed study in which the mechanism is known or can be controlled for. This may not be possible if data are obtained retrospectively from different sources, which may be unavoidable in the analysis of very rare disorders. It is therefore useful to seek models in which estimation of parameters (such as penetrance) does not depend on knowledge of the precise mechanism leading to ascertainment.

Censoring introduces further complications. If ascertainment depends on the number of affected persons observed in a family, then it may be affected by any censoring that prevents observation of a possible case.

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

1.4. Can We Correct Ascertainment Bias?

Correction of ascertainment bias is an important theme in the genetics literature; see in particular Ewens & Shute (1986, 1988a, 1988b) and Thompson (1993). However most such methods depend on knowing the sampling mechanism, and being able to find conditional likelihoods, conditioning on the reasons for ascertainment. A simple thought-experiment shows the difficulty of correcting ascertainment bias in retrospective analyses of family histories, where the reasons for ascertainment might include the observation of a large number of affected members.

(a) Consider a gene in which mutations, which are rare, greatly increase the probability of some disease. Suppose it is a known biological fact that mutations are completely homogeneous in their effects, meaning that a 'true' rate of onset $\mu^*(x)$ of mutation carriers actually exists.

(b) By chance alone, some families in which the mutation is inherited will have an unusually large number of affected members. Geneticists trying to find the gene search the world's medical histories and find these families. They do not sample the other extreme at all, carrier families which by chance have very few affected members.

(c) Rates of onset based on *retrospective* analysis of these families will inevitably be inflated. It is difficult to see how this might be corrected, without knowing the distribution of the mutations in unselected families. Estimates based on *prospective* studies of those family members not affected at the time of the study ought to be unbiased, but this is expensive and time-consuming.

This problem directly affects insurance applications. In most countries, currently, genetic testing is only available in a clinical setting, and may only be offered to people who have a family history. In the case of disorders such as Huntington's disease, which have no known causes except mutations in a specific gene, just one affected blood relative might be sufficient reason for offering genetic testing. But other conditions such as breast cancer are common diseases with rare inherited forms, so there would need to be evidence of quite a 'strong' family history before referral to a genetics clinic. It is sometimes argued, therefore, (see Daykin *et al.* (2003)) that a woman who applies for insurance in the knowledge that she carries a mutation in the BRCA1 or BRCA2 genes that cause breast cancer, must be a member of one of these 'high risk' families, and that risk estimates based on studies of these families are appropriate for use in actuarial calculations. Most major studies in the 1990s were retrospective analyses of high risk families, in particular those carried out by the Breast Cancer Linkage Consortium (Ford *et al.*, 1998), which 'recruited' families from all over the world. However, as shown above, retrospective analysis leaves open the possibility that asymptomatic members of such families are at no more risk than mutation carriers in the general population.

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

1.5. The Aims of This Paper

Gui & Macdonald (2002a) suggested a non-parametric estimator for a certain function of the rate of onset of EOAD associated with mutations in the PSEN-1 gene. The reason for doing so was that EOAD is one of the conditions identified as being relevant for insurance by the Association of British Insurers, but there were no existing studies of ages at onset in the genetics literature.

Implicit in their treatment were the assumptions that $\sigma = 1$ (full penetrance) and $p = 1/2$ (no ascertainment bias). Their estimator, while useful, displayed some pathological behaviour (see Sections 2.3 and 2.5) that they suggested might be caused by the presence of ascertainment bias and/or non-random censoring. The key to this behaviour lies in the observation that if all decrements except onset of the genetic disorder are absent, then the survival probability at very high ages is not 0, as in the life table, but $(1 - p\sigma)$. Given full penetrance and no ascertainment bias, this is just 1/2 (the proportion of non-mutation carriers) but otherwise it depends on the sampling scheme, which may be unknown. The aims of this paper are to extend the estimate in Gui & Macdonald (2002a) to allow for $p \neq 1/2$ and $\sigma \neq 1$, opening the way to apply the modified estimate to some questions of critical illness insurance and life insurance as in Gui & Macdonald (2002b).

In Section 2 we discuss the problems of estimating the rate of onset of a dominant disorder, and describe the estimator used by Gui & Macdonald (2002a), which was a variation of the classical Nelson-Aalen estimate. This estimate has an intrinsic limit related to the sampling mechanism, which immediately suggests how we might adjust the estimator to allow for the value of $p\sigma$ estimated from the data. We find that identifiability is a problem because we cannot estimate p and σ separately, but only their product $p\sigma$. In Section 3, we introduce a model for sampling based on numbers of affected persons in a sibship, leading to ascertainment bias, and show how it affects the survival probability in the presence of censoring. In particular we find that if such a form of ascertainment is applied to censored data, as will usually be the case in practice, then even censoring independent of the event of interest does affect the results. We re-analyse the EOAD data from Gui & Macdonald (2002a) in Section 4, and because of the unidentifiability we obtain not a single estimator but a range. Since the adjustment removes the pathological behaviour noted in Gui & Macdonald (2002a), it is an improvement. Our conclusions are in Section 5.

2. Estimating The Rate of Onset of a Dominant Disorder

2.1. The Classical Nelson-Aalen Estimate Applied to Disease Onset

Suppose we observe a sample of N persons, and record the times at which they suffer onset of a certain disease. Observation may be censored, as is common in survival studies. In this case, death before disease onset would be a type of censoring. We suppose that the rate (or force) of onset is a function $\mu(x)$ of age x, and the problem is to estimate it.

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

The observations of the i^{th} person may be described by the sample paths of two simple stochastic processes:

(a) $\mathbf{N}^i(x)$ is the number of observed cases of onset by age x. Clearly it is 0 or 1.

(b) $\mathbf{Y}^i(x)$ is the indicator of being healthy and under observation at age x, equal to 1 if this is true, or 0 if it is false. In the absence of censoring $\mathbf{Y}^i(x) = 1 - \mathbf{N}^i(x)$, but $\mathbf{Y}^i(x)$ can represent a wide range of censoring schemes.

Denote the aggregated observations by the processes $\mathbf{N}(x) = \sum_{i=1}^{i=N} \mathbf{N}^i(x)$ and $\mathbf{Y}(x) = \sum_{i=1}^{i=N} \mathbf{Y}^i(x)$. Let $t_1, t_2, \ldots$ be the times of the observed cases of onset, and $d\mathbf{N}(t_j)$ is the number of cases of onset observed at time t_j. Then the classical Nelson-Aalen estimate is the sum:

$$\sum_{t_j \le x} \frac{d\mathbf{N}(t_j)}{\mathbf{Y}(t_j)} \tag{4}$$

and it is an estimate of the integrated force of onset (or cumulative hazard) $\int_0^x \mu(t)\,dt$.

In a modern formulation, the Nelson-Aalen estimate would be written as a stochastic integral: define $\mathbf{J}(x) = \mathbf{I}_{\{\mathbf{Y}(x)>0\}}$, with the convention that $\mathbf{J}(x)/\mathbf{Y}(x) = 0$ if $\mathbf{Y}(x) = 0$; then we have:

$$\sum_{t_j \le x} \frac{d\mathbf{N}(t_j)}{\mathbf{Y}(t_j)} = \int_0^x \frac{\mathbf{J}(t)}{\mathbf{Y}(t)}\, d\mathbf{N}(t). \tag{5}$$

In this framework, all the properties of the estimator can be obtained, and it is easily seen how it can be used in any multiple-state model; see Andersen *et al.* (1993) for details. In particular, its variance can be estimated reasonably well by:

$$\int_0^x \frac{\mathbf{J}(t)\big(\mathbf{Y}(t) - \Delta\mathbf{N}(t)\big)}{\big(\mathbf{Y}(t)\big)^3}\, d\mathbf{N}(t). \tag{6}$$

2.2. Identification of Mutation Carriers and Conditioning

If, as in the case of some genetic disorders, only mutation carriers may be affected, how may we identify them?

(a) There might be a reliable genetic test.

(b) If symptoms develop, carrier status may be inferred if the disease has no known cause except a gene mutation, or is so rare that sporadic occurrence within an affected family may be neglected. (This excludes common

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

diseases of which a small proportion is caused by dominantly inherited mutations, such as BC.)

(c) Survival free of symptoms does not rule out the possibility of carrying a mutation, unless the mutation has 100% penetrance before very old ages, which is rare.

It might be imagined that the advent of DNA-based genetic tests means that (b) and (c) above will soon be redundant, but this is not so. The prevalence of genetic testing is rather low when there is no effective treatment for the disorder. For example Meiser & Dunn (2000) estimate the prevalence of testing for HD at only 10-20%. Therefore, the mutation status of the majority of family members included in a study may be unknown.

Given a model of inheritance, such as Mendel's laws, it is simple in principle to write down a likelihood, summing over all possible joint genotypes, weighted by the probabilities of those genotypes given by the model (Elston, 1973). This is the most common approach. When questions of ascertainment arise, however, it is impossible to write down a likelihood without formulating a model of how the families were selected for the sample. Here, we assume that this may unknown, so a fully parametric likelihood method cannot be used, and at best, some kind of semi-parametric model will be needed.

Many approaches to estimating rates of onset (including the Nelson-Aalen estimate) are variations of the simple occurrence/exposure rate. The problem of unknown genotypes has been overcome in the past by weighting each person's exposures by the probability that they are a mutation carrier, conditional on all the observations. (Elandt-Johnson, 1973; Newcombe, 1981; Harper & Newcombe, 1992). However, since such studies are usually retrospective (pedigrees may include several generations) it is necessary to consider what information may legitimately be used for conditioning.

For example, it may be known that person X suffered onset of the disorder when they were age 40, many calendar years before the investigation now taking place. Can we therefore use that to say that they were *known* to be a mutation carrier when we are calculating their contribution to the exposure at age 30 (say)? The estimates in Newcombe (1981), following Elandt-Johnson (1973) did make this assumption. This amounts to using conditional probabilities of the form:

$$\mathrm{P}[\text{Onset at age 30} \mid \text{Known mutation carrier}]. \tag{7}$$

Whatever method is used, probabilities or expectations like these will contribute to the estimating equations. But it was onset of the disorder itself that showed that X carried the mutation, and this event is part of the information structure, if we formulate the problem in a probabilistic model (Section 2.3). Equation (7) should really be:

$$\mathrm{P}[\text{Onset at age 30} \mid \text{Observation of event that revealed carrier status}]$$
$$= \mathrm{P}[\text{Onset at age 30} \mid \text{Onset at age 40}] = 0$$

and the estimating equations collapse.

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

The same problem will arise if the event that reveals the mutation status is a presymptomatic genetic test. To avoid it, we must avoid all such conditioning. See Gui & Macdonald (2002a) for a detailed discussion, including the point that pedigrees sometimes do not include enough information to allow each family member to be tracked through several different risk groups, depending on what was known about their relatives at every age.

This conditioning problem essentially arises from an attempt to see into the future. To avoid it, we turn to stochastic process models adapted to the information available at each age, rather than to all the information available retrospectively. In the non- or semi-parametric case, this leads to the Nelson-Aalen estimate. However, we then have to allow for the mutation status being unknown before onset.

2.3. A Nelson-Aalen Estimate and a Bound

Gui & Macdonald (2002a) proposed the continuous-time Markov model in Figure 1, in respect of a person who is at risk of carrying a mutation in the PSEN-1 gene. 'At risk' means that one of their parents carries a PSEN-1 mutation. They assumed, as in Section 1, that the at-risk child inherited it with probability 1/2. Ignoring genetic tests for now, we assume, more generally, that this person, at birth, was in state 0 with probability p, or in state 1 with probability $(1-p)$. Onset of EOAD is represented by transition into state 2, with transition intensities (rates of onset) $\mu_{02}(x)$ and $\mu_{12}(x)$. The former is the real object of interest and, ultimately, the target for estimation. The latter may be assumed to be zero, because EOAD is very rare (about 15 per 100,000 persons, though this is very uncertain (Gui & Macdonald, 2002b)).

Just as the lifetime penetrance σ in a population selected by a particular sampling method might differ from the 'true' penetrance σ^*, so the intensity

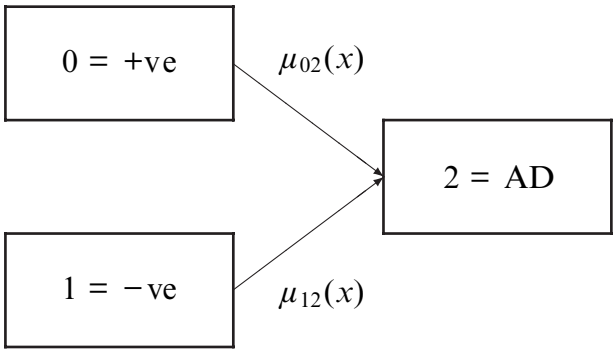

FIGURE 1: A model of the incidence of Alzheimer's disease where an individual may have an EOAD mutation (State 0, +ve) or may not have an EOAD mutation (State 1, –ve).
Source: Gui & Macdonald (2002a).

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

$\mu_{02}(x)$ might depend on the sampling method. When we have cause to refer to the 'true' population intensity we will denote it $\mu^*_{02}(x)$. Again, we emphasise that $\mu_{02}(x)$ is not an empirical (functional) parameter based on any particular sample but a parameter of a model that includes the sampling mechanism.

In respect of the i^{th} life, as in Section 2.1, we record the number of cases of onset, $\mathbf{N}^i(x)$, and the indicator of being at risk, $\mathbf{Y}^i(x)$. Only now, because we cannot distinguish carriers from non-carriers before onset, $\mathbf{Y}^i(x)$ indicates presence in *either* state 0 or state 1, which being unknown, and $\mathbf{N}^i(x)$ represents onset regardless of the originating state. Gui & Macdonald (2002a) showed that the Nelson-Aalen estimate of Equation (5) is then an estimate of the following function of age x (we have changed the notation slightly to emphasise the dependence on p and σ):

$$\Lambda(p,\sigma,x) = \int_0^x \frac{p\exp\left(-\int_0^t \mu_{02}(s)\,ds\right)}{p\exp\left(-\int_0^t \mu_{02}(s)\,ds\right) + (1-p)}\,\mu_{02}(t)\,dt. \tag{8}$$

Strictly, we ought to write $\Lambda(p,\{\mu_{02}(t)\}_{t\le x},x)$, but we allow the scalar parameter σ to remind us of this more concisely. The integrand in Equation (8) is interpreted as the intensity of onset of EOAD in respect of a person who is not known to be a PSEN-1 mutation carrier, but who was born into state 0 with probability p.

In the absence of any decrement other than EOAD, those who do not develop it will live forever. Assuming the lifetime penetrance of PSEN-1 mutations to be σ, the survival probability associated with this hazard, $\exp(-\Lambda(p,\sigma,x))$, tends to $(1-p\sigma)$ instead of to 0, and so $\lim_{x\to\infty}\Lambda(p,\sigma,x) = -\log(1-p\sigma)$. If $p = 1/2$, and $\sigma = 1$, this limit is $\log 2 = 0.693$.

However, a Nelson-Aalen estimate $\hat{\Lambda}(p,\sigma,x)$ is an increasing step function that need not observe any finite limit, especially as the numbers exposed to risk dwindle at higher ages. It can be shown (Section 2.5) that if $\hat{\Lambda}(p,\sigma,x)$ exceeds $-\log(1-p)$, $\hat{\mu}_{02}(x)$ explodes to infinity. Thus for PSEN-1 mutations Gui & Macdonald (2002a), assuming $p = 1/2$ and $\sigma = 1$, found that $\hat{\Lambda}(p,\sigma,x)$ exceeded $\log 2$ by about age 50, and reached about 1.3 by age 60; estimates of $\mu_{02}(x)$ seemed unreliable after about age 45.

2.4. An Example

As an example, following Palamidas (2001) we suppose that $\mu_{02}(x) = 0.285253 - 0.0227997x + 0.0004594x^2$ for $25 \le x \le 60$. This is a hypothetical rate of onset that results in almost 100% penetrance by age 60, obtained by fitting the 'maximum exposure' estimate, males and females combined, in Gui & Macdonald (2002a) (because of missing data, estimates were based on both minimum and maximum exposures to risk (the $\mathbf{Y}(x)$ process) consistent with the data).

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

TABLE 2

A HYPOTHETICAL DISTRIBUTION OF FAMILY SIZES $P[\mathbf{W}_i = w]$.

Number w	1	2	3	4	5	6	7	>7
Probability	0.23	0.5	0.2	0.054	0.012	0.003	0.001	0

We suppose that in a sample of sibships, the size of the i^{th} sibship is a random variable $\mathbf{W}_i$, and that the $\{\mathbf{W}_i\}$ are mutually independent.

(a) $\mathbf{W}_i$ has the distribution in Table 2 (from Macdonald, Waters & Wekwete (2003)).

(b) Each member of each sibship carries a mutation with probability 1/2.

(c) There is no ascertainment bias; even sibships with no affected members are included in the sample.

(d) There is no censoring; every member of each sibship is observed until age 60.

Figure 2 shows the true value of $\Lambda(p,\sigma,x)$, tending to its theoretical limit of log2, and ten simulated examples of its Nelson-Aalen estimate $\hat{\Lambda}(p,\sigma,x)$, each based on a small sample of 25 sibships. Four of these exceed log2. Figure 3 shows 10 simulated estimates $\hat{\Lambda}(p,\sigma,x)$ each based on a very large sample of 10,000 sibships. This shows the convergence to the true $\Lambda(p,\sigma,x)$.

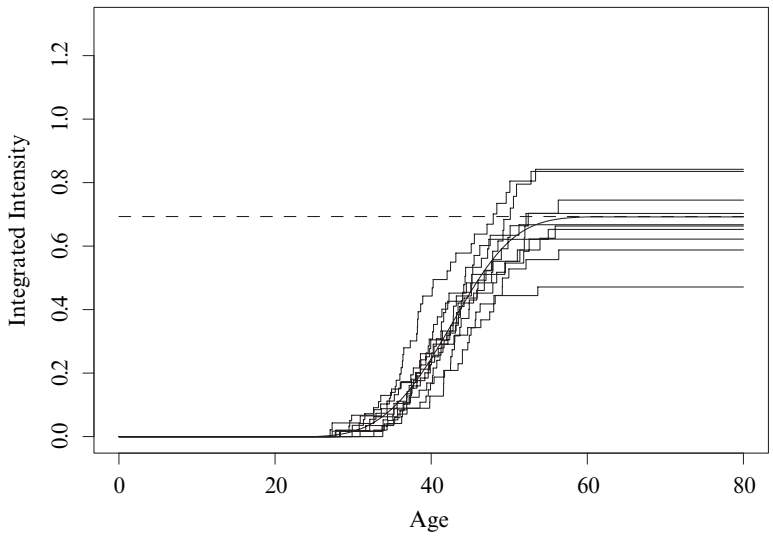

FIGURE 2: The true value of $\Lambda(p,\sigma,x)$ (bold) and 10 simulated Nelson-Aalen estimates of $\Lambda(p,\sigma,x)$ each based on a small sample of 25 sibships. No ascertainment bias or censoring. The theoretical log2 limit is shown by the dotted line.

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

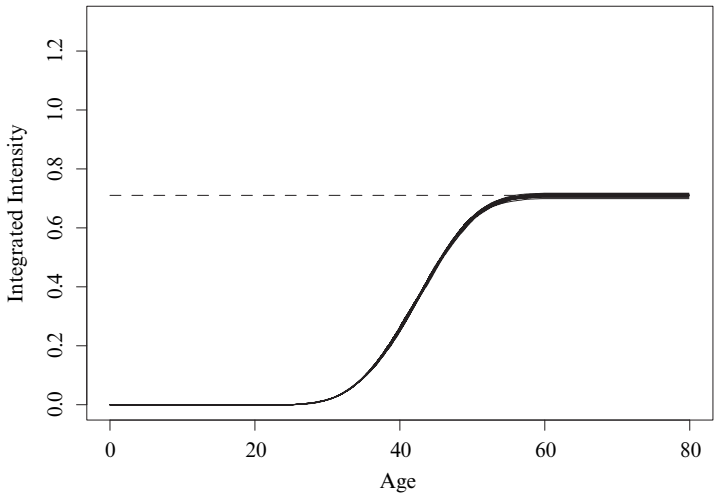

FIGURE 3: The true value of $\Lambda(p,\sigma,x)$ (bold) and 10 simulated Nelson-Aalen estimates of $\Lambda(p,\sigma,x)$ each based on a large sample of 10,000 sibships. No ascertainment bias or censoring. The theoretical log2 limit is shown.

2.5. The Bound on the Nelson-Aalen Estimate

In this section we look in more detail at the bound $-\log(1-p\sigma)$ of the Nelson-Aalen estimate. The real target of estimation is $\mu_{02}(x)$. From Equation (8):

$$\frac{d}{dx}\Lambda(p,\sigma,x) = \frac{p\exp\left(-\int_0^x \mu_{02}(s)\,ds\right)}{p\exp\left(-\int_0^x \mu_{02}(s)\,ds\right)+(1-p)}\,\mu_{02}(x) \tag{9}$$

and if we substitute a smoothed version of $\hat{\Lambda}(p,\sigma,x)$ on the left hand side of this equation, we can solve it numerically for an estimate $\hat{\mu}_{02}(x)$ of $\mu_{02}(x)$. In fact we can express Equation (9) as an ODE:

$$\frac{d}{dx}f(x) + c(x)f(x) = \frac{p-1}{p}c(x) \tag{10}$$

where $c(x) = d\Lambda(p,\sigma,x)/dx$ and $f(x) = \exp\left(-\int_0^x \mu_{02}(t)\,dt\right)$. This ODE has several interesting consequences:

(a) Solving it with $f(0) = 1$ and $\Lambda(p,\sigma,0) = 0$, we get:

$$\exp\left(-\int_0^x \mu_{02}(t)\,dt\right) = \frac{(1-p)^{-1} - e^{\Lambda(p,\sigma,x)}}{p(1-p)^{-1}e^{\Lambda(p,\sigma,x)}} \tag{11}$$

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

which suggests another numerical approach to estimating $\mu_{02}(x)$ from $\hat{\Lambda}(p,\sigma,x)$.

(b) In the limit only non-carriers and unaffected carriers will be left in the population so $\lim_{x\to\infty}\Lambda(p,\sigma,x) = -\log(1-p\sigma)$, which is confirmed by substituting σ on the left hand side of Equation (11) as $x\to\infty$. The intuitive content of this limit was discussed in Section 2.3.

2.6. A Possible Correction for Ascertainment Bias?

At first sight this intrinsic bound on $\Lambda(p,\sigma,x)$ is nothing but a nuisance, curtailing the estimation of $\mu_{02}(x)$ at higher ages. Gui & Macdonald (2002b) had to extrapolate their estimate of $\mu_{02}(x)$ up to age 60 in order to apply it to some insurance problems, and also had to investigate the effect of considerably lower rates of onset (reduced fairly arbitrarily by 50% and 75%) to allow for the possibility of ascertainment bias.

We can see that the bound on $\Lambda(p,\sigma,x)$, $-\log(1-p\sigma)$, is an increasing function of $p\sigma$. This suggests that the Nelson-Aalen estimate $\hat{\Lambda}(p,\sigma,x)$ might exceed its bound with $p\sigma = 1/2$ (namely $\log 2$) not just because of diminishing exposures, but because this is the wrong bound; the assumption that $p\sigma = 1/2$ may be invalid if there is ascertainment bias. However, it also suggests a way to adjust estimates of $\mu_{02}(x)$ for the ascertainment bias.

The adjustment is simple. $\hat{\Lambda}(p,\sigma,x)$ at high ages is taken as an estimate of the limit $-\log(1-p\sigma)$, and hence an estimate of $p\sigma$ is obtained. Here, we have a problem of unidentifiability. The model may be described as semiparametric,

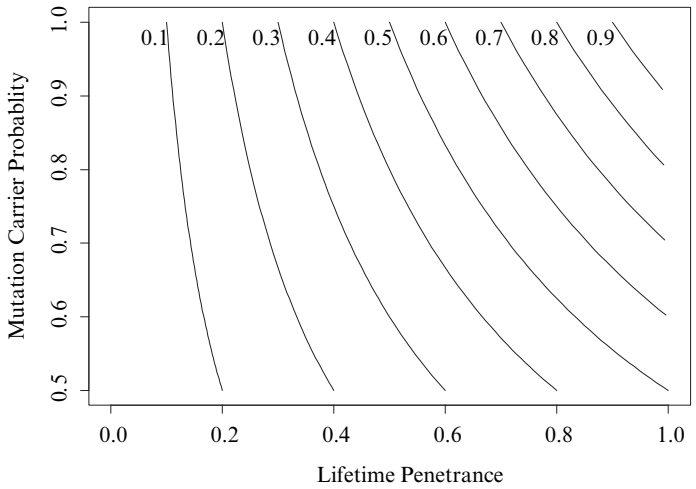

FIGURE 4: Contours of constant $p\sigma$ (values shown at the left of each contour) in a plot of probability p against lifetime penetrance σ. The extreme case of full penetrance ($\sigma = 1$) and no ascertainment bias $p = 1/2$ is at the bottom right.

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

since it involves the two scalar parameters p and σ, but we can only make inferences about the product $p\sigma$, or make suitable *a priori* assumptions about one or other of p and σ.

We still have useful information despite our inability to estimate σ and p separately. Figure 4 shows some contours of constant values of $p\sigma$ in the feasible part of the σ-p plane (with the obvious constraints $1/2 \le p \le 1$ and $0 \le \sigma \le 1$). We see, for example, that a value of $p\sigma \approx 0.7$ as in Section 3.3 is consistent with lifetime penetrance (in the sample) of between 75% and 100%, and with ascertainment bias resulting in p greater than 0.7. For possible values of p consistent with $p\sigma$, we then apply Equation (11) to a smoothed version of $\hat{\Lambda}(p, \sigma, x)$.

We shall find that, despite the unidentifiability problem, we are able to find upper and lower limits for the rate of onset $\mu_{02}(x)$ which improve upon the arbitrary reductions used by Gui & Macdonald (2002b). Before doing that, we need to analyse in more detail how ascertainment might be affected if it is based on censored data.

3. A Model for Ascertainment Bias and Censoring

3.1. A Mechanism for 'Recruiting' Subjects to a Study

Ascertainment bias arises because the subjects 'recruited' to a study of an inherited disorder are often not single individuals, but complete sibships (all the children borne by two parents) or several generations of sibships, or sets of sibships related by having common ancestors. Bias arises if sibships might be selected because of larger numbers of affected members or, more accurately, *observed* affected members. Unless we limit the analysis to completed cohorts, members who are affected after the investigation takes place are not observed to be affected — observation is censored. Censoring could also happen for other reasons, for example losing touch with relatives, or premature death. In this section we will show that censoring taking place before ascertainment affects $p\sigma$, and so is important for our analysis. We propose the following quite general model.

(a) The study is retrospective, taking place at a fixed epoch and obtaining its 'subjects' from historical records of family histories of the disorder (pedigrees).

(b) We sample S^* sibships without bias from the population of affected families. The i^{th} sibship has $\mathbf{W}_i$ members as before.

(c) The number of mutation carriers in the i^{th} sibship is a Binomial$(\mathbf{W}_i, 1/2)$ random variable $\mathbf{M}_i$. We define $\mathbf{M}_{i,j}$ to be the indicator that the j^{th} member of the i^{th} sibship is a mutation carrier.

(d) The number of affected mutation carriers in the i^{th} sibship is a Binomial$(\mathbf{M}_i, \sigma^*)$ random variable $\mathbf{Z}_i$. Note that we use the population penetrance σ^*

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

defined in Equation (3). Conditional on $\mathbf{M}_i$, $\mathbf{Z}_i$ is independent of $\mathbf{W}_i$. We define $\mathbf{Z}_{i,j}$ to be the indicator that the j^{th} member of the i^{th} sibship is an affected mutation carrier

(e) The number of affected members *observed* in the i^{th} sibship is a random variable $\mathbf{X}_i$. In the absence of censoring, $\mathbf{X}_i = \mathbf{Z}_i$, otherwise $X_i \leq \mathbf{Z}_i$. We suppose that censoring is independent of the carrier status of any person. Conditional on $\mathbf{Z}_i$, $\mathbf{X}_i$ is independent of $\mathbf{M}_i$ and $\mathbf{W}_i$. We define $\mathbf{X}_{i,j}$ to be the indicator that the j^{th} member of the i^{th} sibship is observed to be affected.

(f) Sampled sibships are 'selected' for the study by a probabilistic mechanism that makes them unrepresentative of affected families as a whole; of course this mechanism is hidden from the investigator. Define $\mathbf{I}_i = 1$ if the i^{th} sibship sampled is accepted, and $\mathbf{I}_i = 0$ if it is rejected. Conditional on $\mathbf{X}_i$, $\mathbf{I}_i$ is independent of $\mathbf{Z}_i$, $\mathbf{M}_i$ and $\mathbf{W}_i$. Sibships with larger numbers of members observed to be affected are more likely to be accepted. In this model, censoring precedes selection into the study. This seems reasonable for an entirely retrospective study.

(g) Sampling and acceptance/rejection continues until S sibships have been accepted, out of S^* sampled, and $S^* - S$ have been rejected.

Summing over all accepted sibships, define $\mathbf{W} = \sum_{i=1}^{i=S} \mathbf{W}_i$, $\mathbf{M} = \sum_{i=1}^{i=S} \mathbf{M}_i$, $\mathbf{Z} = \sum_{i=1}^{i=S} \mathbf{Z}_i$ and $\mathbf{X} = \sum_{i=1}^{i=S} \mathbf{X}_i$.

3.2. The Effect of Ascertainment Bias and Censoring on the Bound for $\Lambda(p,\sigma,x)$

By the definition of p and σ:

$$p\sigma = \mathrm{P}[\mathbf{Z}_{i,j} = 1 \mid \mathbf{I}_i = 1] \tag{12}$$

$$= \mathrm{P}[\mathbf{Z}_{i,j} = 1, \mathbf{M}_{i,j} = 1 \mid \mathbf{I}_i = 1] \tag{13}$$

$$= \mathrm{P}[\mathbf{M}_{i,j} = 1 \mid \mathbf{I}_i = 1]\, \mathrm{P}[\mathbf{Z}_{i,j} = 1 \mid \mathbf{M}_{i,j} = 1, \mathbf{I}_i = 1]. \tag{14}$$

This shows clearly why we should not assume that $\sigma = \sigma^*$. We have equality if:

$$\mathrm{P}[\mathbf{Z}_{i,j} = 1 \mid \mathbf{M}_{i,j} = 1, \mathbf{I}_i = 1] = \mathrm{P}[\mathbf{Z}_{i,j} = 1 \mid \mathbf{M}_{i,j} = 1] \tag{15}$$

that is, if the ascertainment (following censoring) has no effect on the penetrance. However, the right hand side of Equation (15) is clearly unaffected by censoring, while the left-hand side may be affected by, for example, the study ending when several siblings in the i^{th} sibship are still at risk. So even if the ascertainment mechanism were such that Equation (15) would be true, censoring could still change σ. It is worth repeating that σ is *not* the empirical penetrance based on the sample; it is a parameter, the penetrance in the presence of the

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

ascertainment and censoring mechanism. Equations (12) *et seq.* are not small-sample properties, but lead to asymptotic limits in the sense that as the sample size increases, $\mathbf{Z}/\mathbf{W}$ tends to $p\sigma$, not to $\sigma^*/2$.

Clearly $\mathrm{E}[\mathbf{Z}] = p\sigma\mathrm{E}[\mathbf{W}]$ so, making use of the conditional independences noted in Section 3.1, and allowing for the ascertainment to follow any censoring, we have:

$$p\sigma = \frac{\sum_{w=1}^{w=\infty}\sum_{m=0}^{m=w}\sum_{z=0}^{z=m}\sum_{x=0}^{x=z} z\,\mathrm{P}_{(\mathbf{I}_i|\mathbf{X}_i)}(1|x)\,\mathrm{P}_{(\mathbf{X}_i|\mathbf{Z}_i)}(x|z)\,\mathrm{P}_{(\mathbf{Z}_i|\mathbf{M}_i)}(z|m)\,\mathrm{P}_{(\mathbf{M}_i|\mathbf{W}_i)}(m|w)\,\mathrm{P}_{(\mathbf{W}_i)}(w)}{\sum_{w=1}^{w=\infty}\sum_{m=0}^{m=w}\sum_{z=0}^{z=m}\sum_{x=0}^{x=z} w\,\mathrm{P}_{(\mathbf{I}_i|\mathbf{X}_i)}(1|x)\,\mathrm{P}_{(\mathbf{X}_i|\mathbf{Z}_i)}(x|z)\,\mathrm{P}_{(\mathbf{Z}_i|\mathbf{M}_i)}(z|m)\,\mathrm{P}_{(\mathbf{M}_i|\mathbf{W}_i)}(m|w)\,\mathrm{P}_{(\mathbf{W}_i)}(w)}. \tag{16}$$

We can therefore compute $p\sigma$ if all the necessary conditional distributions are known. This will be unusual, except in hypothetical examples, and even then any realistic censoring will make $\mathrm{P}_{(\mathbf{X}_i|\mathbf{Z}_i)}(x|z)$ intractable. However, as an example, Table 3 shows $p\sigma$ with:

(a) 'true' population penetrance of $\sigma^* = 1$ or $\sigma^* = 0.7$;

(b) a 'true' Mendelian carrier probability of 1/2 in at-risk sibships;

(c) either no censoring, or a crude form of censoring that prevents 50% of affected cases from being observed, regardless of age; and

(d) either no ascertainment bias, or the following simple ascertainment mechanism: sibships with no affected members are rejected, sibships with three or more affected members are accepted, and sibships with one or two affected members are accepted with probabilities 1/3 and 2/3 respectively.

We observe the following:

(a) If there is ascertainment bias, then for any value of w, smaller values of z in the numerator are given smaller weight, and larger values larger weight, so p will increase. This is the basis of assuming that $p > 1/2$ if there is ascertainment bias.

TABLE 3

VALUES OF $p\sigma$ FROM EQUATION (16) WITH/WITHOUT CENSORING AND ASCERTAINMENT BIAS.

$\sigma^* = 1.0$			$\sigma^* = 0.7$		
Censoring	Bias	$p\sigma$	Censoring	Bias	$p\sigma$
No	No	0.500000	No	No	0.350000
No	Yes	0.696562	No	Yes	0.607537
Yes	No	0.500000	Yes	No	0.350000
Yes	Yes	0.698712	Yes	Yes	0.608661

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

(b) One of the qualities of the classical Nelson-Aalen estimate is that its large-sample properties are not affected by the presence of independent censoring. Here, however, the fact that censoring occurs before ascertainment means that it affects the value of $p\sigma$, which does feature in the asymptotic limit of the Nelson-Aalen estimate. In this example, the impact is small, but this is not always so as the more realistic example in Section 3.3 will show. What this means for the interpretation of the estimate is discussed in Section 3.4.

3.3. An Example (Continued)

Extend our hypothetical example from Section 2.4 by implementing the same mechanism for accepting or rejecting sibships as in Table 3 In this case Equation (16) gives $p = 0.696562$, so $-\log(1-p) = 1.19258$ (recall that $\sigma = 1$ here). Figure 5 shows ten simulated Nelson-Aalen estimates of $\Lambda(x)$, each with a small sample of 25 sibships.

Figure 6 shows the effect of censoring as well as ascertainment bias, with a large sample of 10,000 sibships. The censoring takes two forms; independent random censoring with hazard rate 0.025 per annum throughout life (which is quite severe), and censoring at the time of the investigation. Each has an effect, but only when ascertainment follows censoring. If censoring follows ascertainment, the variance but not the limit of $\hat{\Lambda}(p,\sigma,x)$ is affected (not shown).

(a) We emphasise that in retrospective studies, the sampling mechanism will usually be unknown.

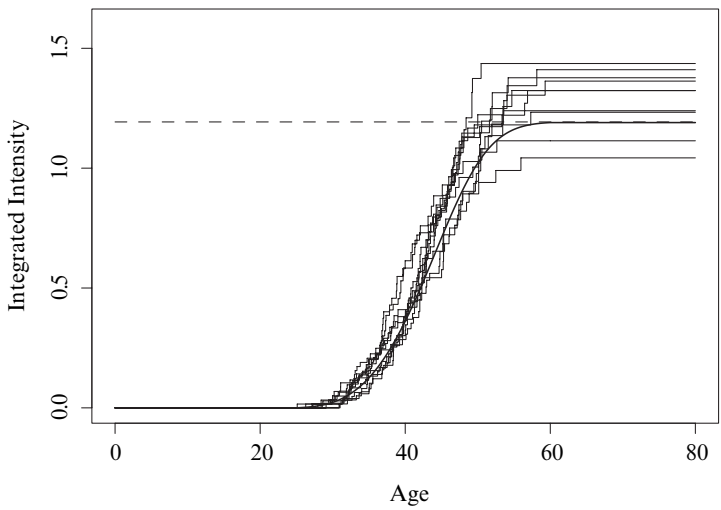

FIGURE 5: True value of $\Lambda(p,\sigma,x)$ (bold) and 10 simulated Nelson-Aalen estimates of $\Lambda(p,\sigma,x)$ each based on a small sample of 25 sibships. Ascertainment bias present but no censoring. The $-\log(1-p)$ limit is shown.

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

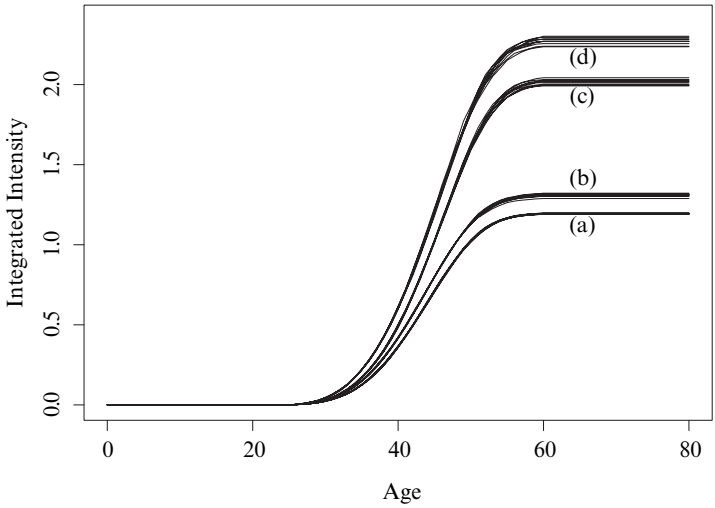

FIGURE 6: The effect of censoring: 10 simulated Nelson-Aalen estimates of $\Lambda(p,\sigma,x)$ each based on a large sample of 10,000 sibships, with ascertainment bias. (a) No censoring. (b) Censoring at time of investigation. (c) Independent censoring at rate 0.025 per annum. (d) Both forms of censoring.

(b) It is encouraging that censoring at the time of the investigation has a much smaller effect than censoring throughout life, since the former will always be present, unless the analysis is limited to completed cohorts, while the latter might be relatively uncommon in carefully researched pedigrees.

(c) Ascertainment following extreme censoring may result in p being close to 1, so the limit of $\hat{\Lambda}(p,\sigma,x)$ may yield a very good estimate of σ. For example in Figure 6, $\hat{\Lambda}(p,\sigma,x)$ reaches about 2.2 with very heavy censoring, implying $\sigma p \approx 0.89$, strong evidence of very high penetrance. Thus, paradoxically, censoring may improve the estimation. The catch is that unfeasibly large samples would be required.

3.4. The Interpretation of the Nelson-Aalen Estimate $\hat{\Lambda}(p,\sigma,x)$

Returning to the model of Figure 1, we see that it is specified by a scalar parameter, the probability p, and a functional parameter, the intensity $\mu_{02}(x)$ (which determines the other scalar parameter, σ). Both p and $\mu_{02}(x)$ are determined by the ascertainment mechanism, which induces dependence upon censoring, even censoring independent of onset, if ascertainment is based on sibships rather than individuals. However what is observable, in any sense, is not p or $\mu_{02}(x)$ but $p\sigma$ or $\Lambda(p,\sigma,x)$.

(a) Adjustment of observations, to obtain an estimate of the 'true' population intensity $\mu^*_{02}(x)$, is intrinsically impossible.

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

(b) However, once $p\sigma$ is estimated we have a range of possible values of p so we can at least locate $\mu_{02}(x)$ within a feasible interval (not to be confused with a confidence interval for $\mu_{02}(x)$; for any given value of p the estimation of confidence intervals for $\mu_{02}(x)$ is a separate exercise). For highly penetrant disorders this may often be sufficient to reach useful conclusions about the insurance implications.

(c) The mechanism of censoring and selection determines $p\sigma$, but with $p\sigma$ given the details of the mechanism disappear from sight, and play no further part in the Nelson-Aalen estimate or its properties. This is why this approach is still useful, because we may have to analyse data retrospectively without any knowledge of what these mechanisms were in the various studies that might have contributed to the data.

4. Application to Presenilin-1 Mutations

4.1. Numbers of Persons At Risk

Gui & Macdonald (2002a) surveyed the literature on PSEN-1 mutations. Because most of these are point mutations, many of them observed in only a single family, the literature includes a fairly large number of published pedigrees, from which estimates can be constructed. In total 47 pedigrees, from over 100 studies, were reported in enough detail to be useable. Even so, certain items of information were often missing:

(a) The Nelson-Aalen estimate assumes that it is known, of each person in the sample, that one of their parents carried the mutation. This may be because the parent or one of the person's siblings has developed EOAD or has had a genetic test. It follows that each person is excluded from the sample until the age at which that information is revealed (this loses some information, but it is required to ensure that the estimate is adapted to the available information).

(b) The age at which observation of unaffected siblings is censored is sometimes omitted. Often the best that can be done is to estimate the highest and lowest possible age at censoring, from information on other relatives, thus minimum and maximum possible exposures. Gui & Macdonald (2002a) estimated rates of onset based on such minimum and maximum possible exposures, which are shown graphically in Figure 9 of Gui & Macdonald (2002a).

(c) The gender of unaffected persons is sometimes omitted.

We refer to Gui & Macdonald (2002a) for details. We take from that study the numbers of observed cases of onset, and the maximum and minimum numbers at risk at each age.

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

4.2. Estimates Adjusted for Ascertainment Bias

Figure 7 shows, at the top, the Nelson-Aalen estimates $\hat{\Lambda}(p,\sigma,x)$, kernel-smoothed versions of these, and approximate 95% confidence intervals (Equation (6)) based on these maximum and minimum exposures.

(a) The maximum and minimum exposures make only a small difference to the estimates.

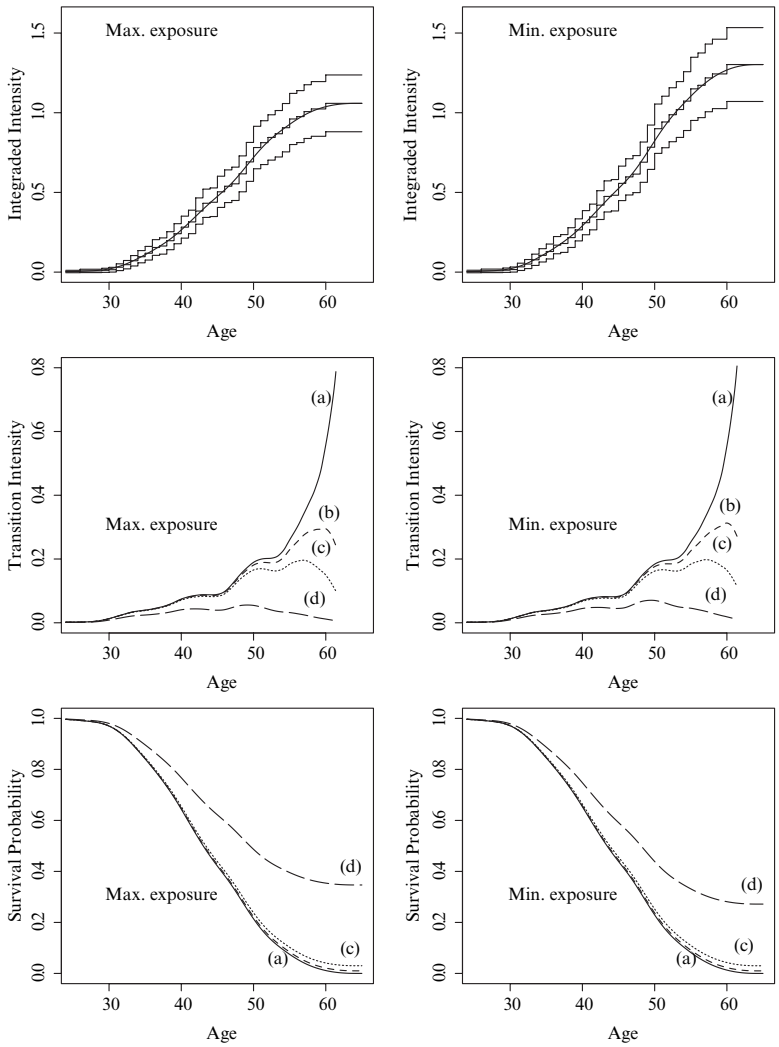

FIGURE 7: Onset rates of EOAD associated with PSEN-1 mutations, with maximum (left) and minimum (right) exposures. Nelson-Aalen estimates $\hat{\Lambda}(p,\sigma,x)$ and 95% CIs (top), estimated intensities $\hat{\mu}_{02}(x)$ (middle) and corresponding survival functions $\exp\left(-\int_0^x \hat{\mu}_{02}(t)dt\right)$ (bottom). Assumed values of penetrance σ are (a) 1; (b) 0.9; (c) 0.8; and (d) 0.653 (maximum exposures) and 0.728 (minimum exposures).

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

(b) The confidence intervals are quite narrow, compared with those usually seen for non-parametric estimates at high ages. This is because the exposures include non-mutation carriers who will almost certainly not develop EOAD, and who therefore remain in the risk set until they are censored.

(c) We take the value of $\hat{\Lambda}(p,65)$ be an estimate of $-\log(1-p\sigma)$, hence $1-\exp(-\hat{\Lambda}(p,65))$ to be an estimate $\widehat{p\sigma}$ of $p\sigma$. The data included a case of AD at age 68, but age 65 is usually taken to be the limit of early-onset cases. With maximum exposures, $\widehat{p\sigma} = 0.653$ (confidence interval (0.587, 0.718)) and with minimum exposures $\widehat{p\sigma} = 0.728$ (0.646, 0.810).

Figure 7 (middle) shows estimates $\hat{\mu}_{02}(x)$ of the intensity of onset, based on a selection of possible values of the penetrance σ consistent with the estimates $\widehat{p\sigma}$. These include the limiting cases of $\sigma = 1$ (full penetrance, labelled (a) in Figure 7) and $\sigma = \widehat{p\sigma}$ (representing such extreme ascertainment bias that $p = 1$, labelled (d) in Figure 7). The corresponding survival functions are shown at the bottom of Figure 7.

(a) The value of σ (equivalently, p) makes a very great difference to the estimated intensity, whereas the difference between maximum and minimum exposures does not. We can safely conclude that the results are fairly robust to the missing data described in Section 4.1.

(b) For any supposed value σ, we estimate p to be $\sigma = \widehat{p\sigma}/\sigma$. With $\sigma = 1$ (full penetrance) that means that the sample contains persons who are mutation carriers with 65.3% probability (maximum exposures) or 72.8% probability (minimum exposure) instead of the 50% probability that we would expect in the absence of ascertainment bias.

4.3. Comparison with Previous Estimates

Figure 8 compares our estimates of the intensity of onset and associated survival functions with those of Gui & Macdonald (2002a). Recall that the latter were obtained with the same Nelson-Aalen estimate as we have used, but without any explicit allowance for ascertainment bias, effectively assuming $p = 1/2$. As shown, the estimates blew up just beyond age 50, though they seemed reasonably well-behaved until the mid-40s. However even the highest of our estimates, with $\sigma = 1$, are considerably lower. Because the intensities are all high, however, the differences between the associated survival functions are much less dramatic. Indeed at the other extreme, our estimates with $p = 1$ (extreme ascertainment bias) imply that fewer than 40% of mutation carriers would escape onset by age 60 (as could be inferred from the estimates $\widehat{p\sigma}$) so the significance for insurance is not diminished.

For insurance modelling, Gui & Macdonald (2002b) recognised that their estimates would be too high, and:

(a) smoothed the lower of their two estimates (that based on maximum exposures);

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

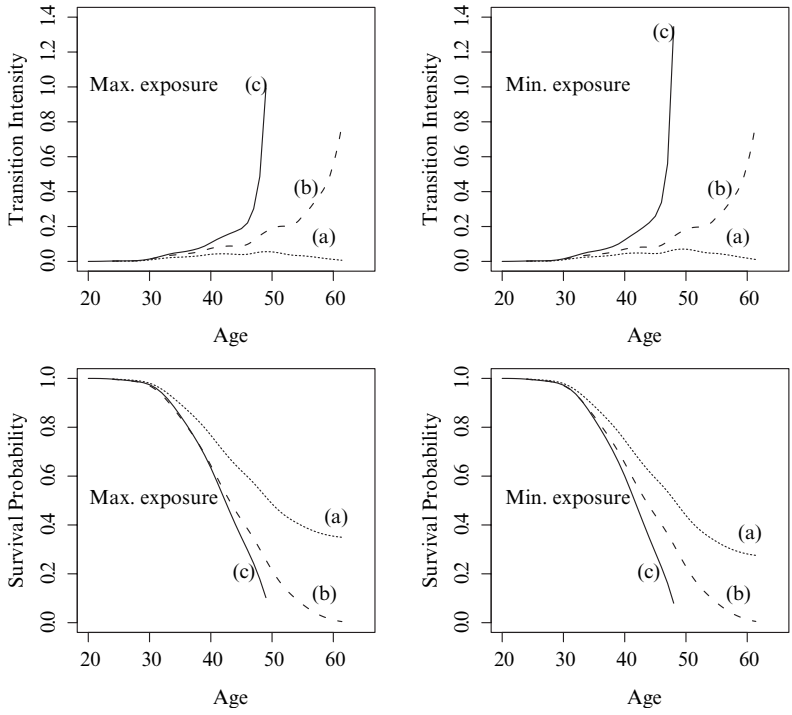

FIGURE 8: Comparison of estimated intensities $\hat{\mu}_{02}(x)$ (top) and survival functions $\exp\left(-\int_0^x \hat{\mu}_{02}(t)dt\right)$ (bottom) with those of Gui & Macdonald (2002a). (a) is our estimate with the lowest possible value of σ; (b) is our estimate with $\sigma = 1$; (c) is from Gui & Macdonald (2002a).

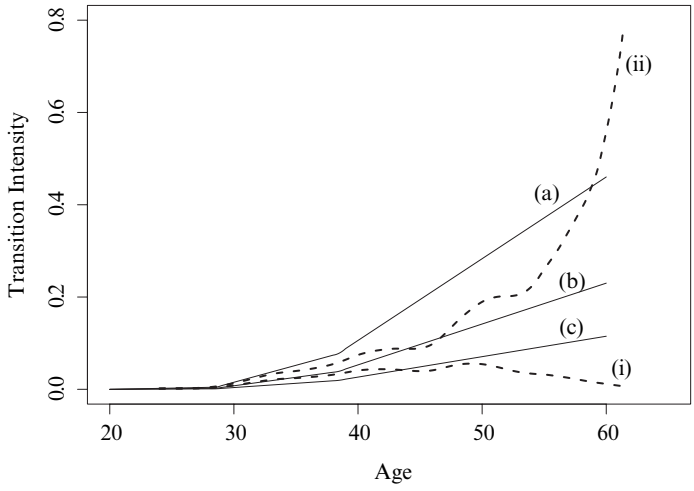

FIGURE 9: Comparison of estimated intensities (maximum exposures) with those used by Gui & Macdonald (2002b). (i) is our estimate with the lowest possible value of σ; (ii) is our estimate with $\sigma = 1$; (a) is the smoothed intensity from Gui & Macdonald (2002b); (b) is 50% of (a); and (c) is 25% of (a).

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

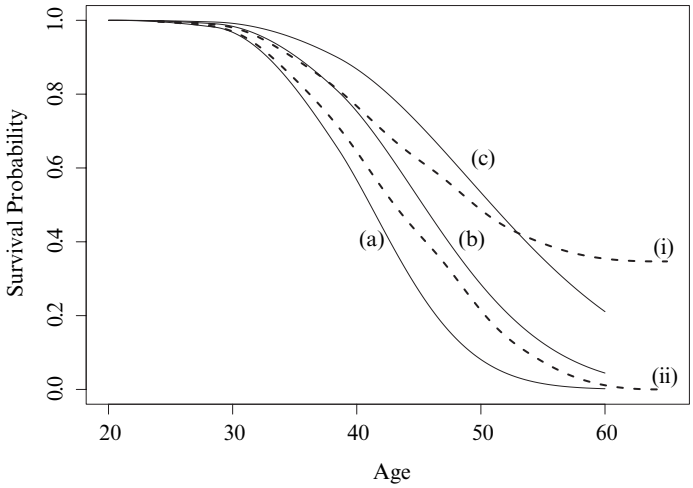

FIGURE 10: Comparison of estimated intensities with those used by Gui & Macdonald (2002b). Based on the intensities in Figure 9, see there for the labels.

(b) extrapolated it to age 60; and

(c) as an *ad hoc* allowance for ascertainment bias, considered intensities that were 50% and 25% of the smoothed estimate.

The resulting intensities are shown in Figure 9, as well as the highest ($\sigma = 1$) and lowest ($p = 1$) of our estimates based on maximum exposures. These suggest that the *ad hoc* reductions made by Gui & Macdonald (2002b) were quite reasonable; perhaps a little low at younger ages but consistent with the range of our estimates above about age 45. Figure 10 shows the corresponding survival functions.

However, the problem remains that we have estimated $p\sigma$ and not $p^*\sigma^*$, and that the latter is intrinsically unobtainable from a retrospective analysis.

5. CONCLUSIONS

Because EOAD is a very rare disease, the selection of the sample of families is unlikely to be by random ascertainment. Families with large numbers of affected members may be more likely to be detected by researchers. Therefore, we expect that the PSEN-1 data were not randomly ascertained. Also, the sampling penetrance related to the PSEN-1 data is not certain and has to be estimated either with or without ascertainment bias.

We concluded that even classical independent censoring has an effect on the integrated intensity estimates when ascertainment follows after censoring. Then, both p and σ are determined by the ascertainment scheme and the censoring.

https://doi.org/10.2143/AST.37.2.2024075 Published online by Cambridge University Press

We extended Gui & Macdonald's (2002a) estimator by introducing the in-sample parameters p and σ. Our estimate, which is a variation of the Nelson-Aalen estimator, has an intrinsic limit related to the sampling mechanism, which immediately suggests how we might estimate the product $p\sigma$. We found that unidentifiability is a problem because we cannot estimate p and σ separately, but only their product $p\sigma$.

An estimate of the 'true' population intensity $\mu^*_{02}(x)$, is intrinsically impossible. However, the range of possible values of (p,σ) allow us to at least locate $\mu_{02}(x)$ within a feasible interval, hence removing the pathological behaviour noted in Gui & Macdonald (2002a).

Acknowledgements

We thank Aikaterina Berou and Athanasios Palamidas, whose MSc theses suggested the line this research should take, and Eng Hock Gui for providing the EOAD data. 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. One of us (CE) was funded during the research by CONACYT.

References

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

Daykin, C.D., Akers, D.A., Macdonald, A.S., McGleenan, T., Paul, D. and Turvey, P.J. (2003) Genetics and insurance – some social policy issues (with discussions), *British Actuarial Journal*, **9**, 787-874.

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.

Elston, R.C. (1973) Ascertainment and age at onset in pedigree analysis, *Human Heridity*, **23**, 105-112.

Ewens, W.J. and Shute, N.C.E. (1986) A resolution of the ascertainment sampling problem I: Theory, *Theoretical Population Biology*, **30**, 388-412.

Ewens, W.J. and Shute, N.C.E. (1988a) A resolution of the ascertainment sampling problem II: Generalizations and numerical results, *American Journal of Human Genetics*, **43**, 374-386.

Ewens, W.J. and Shute, N.C.E. (1988b) A resolution of the ascertainment sampling problem III: Pedigrees, *American Journal of Human Genetics*, **43**, 387-395.

Ford, D., Easton, D.F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D.T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M.D., Struewing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T.R., Tonin, P., Neuhausen, S., Barkardottir, R., Eyfjord, J., Lynch, H., Ponder, B.A.J., Gayther, S.A., Birch, J.M., Lindblom, A., Stoppa-Lyonnet, D., Bignon, Y., Borg, A., Hamann, U., Haites, N., Scott, R.J., Maugard, C.M., Vasen, H., Seitz, S., Cannon-Albright, L.A., Schofield, A., Zelada-Hedman, M. and the Breast Cancer Linkage Consortium (1998) Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families, *American Journal of Human Genetics*, **62**, 676-689.

Gui, E.H. and Macdonald, A.S. (2002a) A Nelson-Aalen estimate of the incidence rates of early-onset Alzheimer's disease associated with the Presenilin-1 gene, *ASTIN Bulletin*, **32**, 1-42.

GUI, E.H. and MACDONALD, A.S. (2002b) *Early-onset Alzheimer's disease, critical illness insurance and life insurance*, Research Report No. 02/2, Genetics and Insurance Research Centre, Heriot-Watt University, Edinburgh.

HARPER, P.S. and 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.

MACDONALD, A.S., WATERS, H.R. and WEKWETE, C.T. (2003) The genetics of breast and ovarian cancer I: A model of family history, *Scandinavian Actuarial Journal*, **2003**, 1-27.

MEISER, B. and DUNN, S. (2000) Psychological impact of genetic testing for Huntington's disease: an update of the literature, *J. Neurol. Neurosurg. Psychiatry*, **69**, 574-578.

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

PALAMIDAS, A. (2001) *Ascertainment bias in genetic epidemiology*. M.Sc. dissertation, Heriot-Watt University, Edinburgh.

SHAM, P. (1998) *Statistics in Human Genetics*. Arnold, London.

THOMPSON, E. (1993) *Sampling and ascertainment in genetic epidemiology: A tutorial review*, Technical Report 243, Department of Statistics, University of Washington.

ANGUS MACDONALD
Department of Actuarial Mathematics and Statistics
and the Maxwell Institute for Mathematical Sciences
Heriot-Watt University, Edinburgh EH14 4AS,
United Kingdom
Tel: +44(0)131-451-3209
Fax: +44(0)131-451-3249
E-mail: A.S.Macdonald@ma.hw.ac.uk